(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 October 2006 (19.10.2006)

PCT

(10) International Publication Number
WO 2006/110478 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2006/012878

(22) International Filing Date: 7 April 2006 (07.04.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/670,061 11 April 2005 (11.04.2005) US

(71) Applicant *(for all designated States except US)*: **NOVARTIS AG** [CH/CH]; Lichtstrasse 35, CH-4056 Basel (CH).

(71) Applicant *(for AT only)*: **NOVARTIS PHARMA GmbH** [AT/AT]; Brunner Strasse 59, A-1230 Vienna (AT).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CULVER, Kenneth, W.** [US/US]; 8 Jared Drive, Mendham, NJ 07945 (US). **ZHU, Jian** [CN/US]; 25 Dunnerdale Road, Morris Plains, NJ 07950 (US). **LILLEBERG, Stan** [US/US]; 2121 S 84th Street, Omaha, NE 68124 (US).

(74) Agent: **PRINCE, John, T.**; NOVARTIS AG, Corporate Intellectual Property, Lichtstrasse 35, CH-4056 Basel (CH).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MUTATIONS AND POLYMORPHISMS OF EPIDERMAL GROWTH FACTOR RECEPTOR

(57) Abstract: This invention relates generally to the analytical testing of tissue samples in vitro, and more particularly to aspects of genetic polymorphisms and mutations of the epidermal growth factor receptor. The invention provides new EGFR mutations and SNPs, useful in the diagnosis and treatment of subjects in need thereof. Accordingly, the various aspects of the present invention relate to polynucleotides encoding the EGFR mutations of the invention, expression vectors encoding the EGFR mutant polypeptides of the invention and organisms that express the EGFR mutant and polymorphic polynucleotides and/or EGFR mutant/polymorphic polypeptides of the invention. The various aspects of the present invention further relate to diagnostic/theranostic methods and kits that use the EGFR mutations and polymorphisms of the invention to identify individuals predisposed to disease or to classify individuals with regard to drug responsiveness, side effects, or optimal drug dose.

WO 2006/110478 A2

MUTATIONS AND POLYMORPHISMS OF EPIDERMAL GROWTH FACTOR RECEPTOR

FIELD OF THE INVENTION

[01] This invention relates generally to the analytical testing of tissue samples *in vitro*, and more particularly to aspects of genetic mutations and polymorphisms of epidermal growth factor receptor.

BACKGROUND OF THE INVENTION

[02] Conventional medical approaches to diagnosis and treatment of disease is based on clinical data alone, or made in conjunction with a diagnostic test. Such traditional practices often lead to therapeutic choices that are not optimal for the efficacy of the prescribed drug therapy or to minimize the likelihood of side effects for an individual subject. Therapy specific diagnostics (*a.k.a.*, theranostics) is an emerging medical technology field, which provides tests useful to diagnose a disease, choose the correct treatment regime and monitor a subject's response. That is, theranostics are useful to predict and assess drug response in individual subjects, *i.e.*, individualized medicine. Theranostic tests are also useful to select subjects for treatments that are particularly likely to benefit from the treatment or to provide an early and objective indication of treatment efficacy in individual subjects, so that the treatment can be altered with a minimum of delay. Theranostics are useful in clinical diagnosis and management of a variety of diseases and disorders, which include, but are not limited to, *e.g.*, cardiovascular disease, cancer, infectious diseases, Alzheimer's disease and the prediction of drug toxicity or drug resistance. Theranostic tests may be developed in any suitable diagnostic testing format, which include, but is not limited to, *e.g.*, immunohistochemical tests, clinical chemistry, immunoassay, cell-based technologies, and nucleic acid tests.

[03] Progress in pharmacogenomics and pharmacogenetics, which establishes correlations between responses to specific drugs and the genetic profile of individual patients and/or their tumours, is foundational to the development of new theranostic approaches. As such, there is a need in the art for the evaluation of patient-to-patient variations and tumour mutations in gene sequence and gene expression. A common form of genetic profiling relies on the identification of DNA sequence variations called single nucleotide polymorphisms ("SNPs"),

which are one type of genetic alteration leading to patient-to-patient variation in individual drug response. In addition, it is well established in the art that acquired DNA changes (mutations) are responsible, alone or in part, for pathological processes. It follows that, there is a need art to identify and characterize genetic mutations and SNPs, which are useful to identify the genotypes of subjects and their tumours associated with drug responsiveness, side effects, or optimal dose.

[04] A key driver for cell growth is the epidermal growth factor (EGF) and the receptor for EGF ("EGFR"). EGFR plays an important role in cellular proliferation as well as apoptosis, angiogenesis and metastatic spread, processes that are crucial to tumour progression. Indeed, studies have shown that EGFR-mediated cell growth is increased in a variety of solid tumours including non-small cell lung cancer, prostate cancer, breast cancer, gastric cancer, and tumours of the head and neck. Salomon DS *et al.*, *Critical Reviews in Oncology/Haematology*, 19:183-232 (1995). Furthermore, excessive activation of EGFR on the cancer cell surface is now known to be associated with advanced disease, the development of a metastatic phenotype and a poor prognosis in cancer patients.

[05] Some mutations of EGFR have been identified in human cancer patients that affect their response to chemotherapy directed toward EGFR. For example, studies by Lynch and co-workers (*N. Engl. J. Med.* 350: 2129-2139 (2004)) and Paez and co-workers (*Science* 304: 1497-1500 (2004)) have described somatic mutations in the EGFR gene in patients with non-small cell lung cancer (NSCLC) who were particularly responsive to an EGFR kinase inhibitor. EGFR mutations have also been identified in patients experiencing responses to the EGFR tyrosine kinase inhibitor erlotinib. Shepherd *et al.*, *Proc. Am. Soc. Clin. Oncol.*, 23: 18 (abstr. # 7022) (2004).

[06] However, the effect of EGFR mutations and their clinical significance is not well understood. Studies are needed to evaluate the role of mutations, as well as other markers, in predicting outcomes of cancer treatment, *e.g.*, NSCLC. Accordingly, there is a need in the art for additional information about the relationship between EGFR mutations and cancer.

SUMMARY OF THE INVENTION

[07] The invention provides for the use of an EGFR modulating agent in the manufacture of a medicament for the treatment of cancer in a selected patient population. The patient population is selected on the basis of the genotype of the patients at an EGFR genetic locus

indicative of efficacy of the EGFR modulating agent in treating cancer. In several embodiments, the cancer can be glioblastoma; melanoma; ovarian cancer; breast cancer; cholangioma; non-small-cell lung cancer (NSCLC); prostate cancer; colon cancer; or myeloma.

[08] The invention also provides an isolated polynucleotide having a sequence encoding an EGFR mutation. In several embodiments, the EGFR mutations are the previously-unidentified mutations listed in TABLE 1. Accordingly, the invention provides vectors and organisms containing the EGFR mutations of the invention and polypeptides encoded by polynucleotides containing the EGFR mutations of the invention.

[09] The invention further provides a method for treating cancer in a subject. The genotype or haplotype of a subject is obtained at an EGFR gene locus, so that the genotype and/or haplotype is indicative of a propensity of the cancer to respond to the drug. Then, an anti-cancer therapy is administered to the subject.

[10] The invention provides a method for diagnosing cancer in a subject and a method for choosing subjects for inclusion in a clinical trial for determining efficacy of an EGFR modulating agent; in both these methods the genotype and/or haplotype of a subject is interrogated at an EGFR gene locus. Also provided by the invention are kits for use in determining a treatment strategy for cancer.

[11] The invention also provides for the use of each of the mutations of the inventions as a drug target.

BRIEF DESCRIPTION OF THE DRAWINGS

[12] The drawing figures depict preferred embodiments by way of example, not by way of limitations.

[13] FIG. 1 is a schematic drawing of the three-dimensional structure of the protein kinase domain of the wild-type EGFR with the location of selected EGFR missense mutation highlighted by an arrow.

[14] FIG. 2 is a chart containing information about previously-known mutations in the EGFR gene.

[15] FIG. 3 is a chart containing information about previously-known and newly-identified mutations in the EGFR gene.

[16] FIG. 4 is a chart containing information about newly-unidentified mutations in the EGFR gene.

DETAILED DESCRIPTION OF THE INVENTION

[17] It is to be appreciated that certain aspects, modes, embodiments, variations and features of the invention are described below in various levels of detail in order to provide a substantial understanding of the present invention. In general, such disclosure provides new epidermal growth factor receptor ("EGFR") mutations and SNPs that may be useful, alone or in combination, in the diagnosis and treatment of subjects in need thereof. Accordingly, the various aspects of the present invention relate to polynucleotides encoding EGFR mutations and polymorphisms of the invention, expression vectors encoding the EGFR mutant polypeptides of the invention and organisms that express the EGFR mutant/polymorphic polynucleotides and/or EGFR mutant/polymorphic polypeptides of the invention. The various aspects of the present invention further relate to diagnostic/theranostic methods and kits that use the EGFR mutations and/or polymorphisms of the invention to identify individuals predisposed to disease or to classify individuals and tumours with regard to drug responsiveness, side effects, or optimal drug dose. In other aspects, the invention provides methods for compound validation and a computer system for storing and analyzing data related to the EGFR mutations and polymorphisms of the invention. Accordingly, various particular embodiments that illustrate these aspects follow.

[18] *Definitions.* The definitions of certain terms as used in this specification are provided below. Definitions of other terms may be found in the glossary provided by the U.S. Department of Energy, Office of Science, Human Genome Project (http://www.ornl.gov/sci/techresources/Human_Genome/glossary/).

[19] As used herein, the term "allele" means a particular form of a gene or DNA sequence at a specific chromosomal location (locus).

[20] As used herein, the term "antibody" includes, but is not limited to, polyclonal antibodies, monoclonal antibodies, humanized or chimeric antibodies and biologically functional antibody fragments sufficient for binding of the antibody fragment to the protein.

[21] As used herein, the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (*i.e.*, side effects).

[22] As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enrol subjects.

[23] As used herein, the term "effective amount" of a compound is a quantity sufficient to achieve a desired pharmacodynamic, toxicologic, therapeutic and/or prophylactic effect, for example, an amount which results in the prevention of or a decrease in the symptoms associated with a disease that is being treated, *e.g.*, the diseases associated with EGFR mutant polypeptides and EGFR mutant polynucleotides identified herein. The amount of compound administered to the subject will depend on the type and severity of the disease and on the characteristics of the individual, such as general health, age, sex, body weight and tolerance to drugs. It will also depend on the degree, severity and type of disease. The skilled artisan will be able to determine appropriate dosages depending on these and other factors. Typically, an effective amount of the compounds of the present invention, sufficient for achieving a therapeutic or prophylactic effect, range from about 0.000001 mg per kilogram body weight per day to about 10,000 mg per kilogram body weight per day. Preferably, the dosage ranges are from about 0.0001 mg per kilogram body weight per day to about 100 mg per kilogram body weight per day. The compounds of the present invention can also be administered in combination with each other, or with one or more additional therapeutic compounds.

[24] Glivec® (Gleevec®; imatinib) is a medication for chronic myeloid leukaemia (CML) and certain stages of gastrointestinal stromal tumours (GIST). It targets and interferes with the molecular abnormalities that drive the growth of cancer cells. Corless CL *et al., J. Clin. Oncol.* 22(18):3813-25 (September 15, 2004); Verweij J *et al., Lancet* 364(9440):1127-34 (September 25, 2004); Kantarjian HM *et al., Blood* 104(7):1979-88 (October 1, 2004). By inhibiting multiple targets, Glivec® has potential as an anticancer therapy for several types of cancer, including leukaemia and solid tumours.

[25] The aromatase inhibitor FEMARA® is a treatment for advanced breast cancer in postmenopausal women. It blocks the use of oestrogen by certain types of breast cancer that require oestrogen to grow. Janicke F, *Breast* 13 Suppl 1:S10-8 (December 2004); Mouridsen H *et al., Oncologist* 9(5):489-96 (2004).

[26] Sandostatin® LAR® is used to treat patients with acromegaly and to control symptoms, such as severe diarrhoea and flushing, in patients with functional gastro-entero-

pancreatic (GEP) tumours (*e.g.* metastatic carcinoid tumours and vasoactive intestinal peptide-secreting tumours [VIPomas]). Oberg K, *Chemotherapy* 47 Suppl 2:40-53 (2001); Raderer M *et al., Oncology* 60(2):141-5 (2001); Aparicio T *et al., Eur. J. Cancer* 37(8):1014-9 (May 2001). Sandostatin® LAR® regulates hormones in the body to help manage diseases and their symptoms.

[27] ZOMETA® is a treatment for hypocalcaemia of malignancy (HCM)1 and for the treatment of bone metastases across a broad range of tumour types. These tumours include multiple myeloma, prostrate cancer, breast cancer, lung cancer, renal cancer and other solid tumours. Rosen LS *et al., Cancer* 100(12):2613-21 (June 15, 2004).

[28] Vatalanib (1-[4-chloroanilino]-4-[4-pyridylmethyl] phthalazine succinate) is a multi-VEGF receptor (VEGFR) inhibitor that may block the creation of new blood vessels to prevent tumour growth. This compound inhibits all known VEGF receptor tyrosine kinases, blocking angiogenesis and lymphangiogenesis. Drevs J *et al., Cancer Res.* 60:4819-4824 (2000); Wood JM *et al., Cancer Res.* 60:2178-2189 (2000). Vatalanib is being studied in two large, multinational, randomized, phase III, placebo-controlled trials in combination with FOLFOX-4 in first-line and second-line treatment of patients with metastatic colorectal cancer. Thomas A *et al., 37th Annual Meeting of the American Society of Clinical Oncology*, San Francisco, CA, Abstract 279 (May 12-15, 2001).

[29] The orally bioavailable rapamycin derivative everolimus inhibits oncogenic signalling in tumour cells. By blocking the mammalian target of rapamycin (mTOR)-mediated signalling, everolimus exhibits broad antiproliferative activity in tumour cell lines and animal models of cancer. Boulay A *et al., Cancer Res.* 64:252-261 (2004). In preclinical studies, everolimus also potently inhibited the proliferation of human umbilical vein endothelial cells directly indicating an involvement in angiogenesis. By blocking tumour cell proliferation and angiogenesis, everolimus may provide a clinical benefit to patients with cancer. Everolimus is being investigated for its antitumour properties in a number of clinical studies in patients with haematological and solid tumours. Huang S & Houghton PJ, *Curr. Opin. Investig. Drugs* 3:295-304 (2002).

[30] Gimatecan is a novel oral inhibitor of topoisomerase I (topo I). Gimatecan blocks cell division in cells that divide rapidly, such as cancer cells, which activates apoptosis. Preclinical data indicate that gimatecan is not a substrate for multidrug resistance pumps, and that it increases the drug-target interaction. De Cesare M *et al., Cancer Res.* 61:7189-7195 (2001).

Phase I clinical studies indicate that the dose-limiting toxicity of gimatecan is myelosuppression.

[31] Patupilone is a microtubule stabilizer. Altmann K-H, *Curr. Opin. Chem. Biol.* 5:424-431 (2001); Altmann K-H *et al., Biochim Biophys Acta.* 470:M79-M91 (2000); O'Neill V *et al., 36th Annual Meeting of the American Society of Clinical Oncology*; May 19-23, 2000; New Orleans, LA, Abstract 829; Calvert PM *et al. Proceedings of the 11th National Cancer Institute-European Organization for Research and Treatment of Cancer/American Association for Cancer Research Symposium on New Drugs in Cancer Therapy*; November 7-10, 2000; Amsterdam, The Netherlands, Abstract 575. Patupilone blocked mitosis and induced apoptosis greater than the frequently used anticancer drug paclitaxel. Also, patupilone retained full activity against human cancer cells that were resistant to paclitaxel and other chemotherapeutic agents.

[32] Midostaurin is an inhibitor of multiple signalling proteins. By targeting specific receptor tyrosine kinases and components of several signal transduction pathways, midostaurin impacts several targets involved in cell growth (*e.g.*, KIT, PDGFR, PKC), leukaemic cell proliferation (*e.g.*, FLT3), and angiogenesis (*e.g.*, VEGFR2). Weisberg E *et al. Cancer Cell* 1:433-443 (2002); Fabbro D *et al., Anticancer Drug Des.* 15:17-28 (2000). In preclinical studies, midostaurin showed broad antiproliferative activity against various tumour cell lines, including those that were resistant to several other chemotherapeutic agents.

[33] The somatostatin analogue pasireotide is a stable cyclohexapeptide with broad somatotropin release inhibiting factor (SRIF) receptor binding. Bruns C *et al., Eur. J. Endocrinol.* 146(5):707-16 (May 2002); Weckbecker G *et al., Endocrinology* 143(10):4123-30 (October 2002); Oberg K, *Chemotherapy* 47 Suppl 2:40-53 (2001).

[34] LBH589 is a histone deacetylase (HDAC) inhibitor. By blocking the deacetylase activity of HDAC, HDAC inhibitors activate gene transcription of critical genes that cause apoptosis (programmed cell death). By triggering apoptosis, LBH589 induces growth inhibition and regression in tumour cell lines. LBH589 is being tested in phase I clinical trials as an anticancer agent. See also, George P *et al., Blood* 105(4):1768-76 (February 15, 2005).

[35] AEE788 inhibits multiple receptor tyrosine kinases including EGFR, HER2, and VEGFR, which stimulate tumour cell growth and angiogenesis. Traxler P *et al., Cancer Res.* 64:4931-4941 (2004). In preclinical studies, AEE788 showed high target specificity and demonstrated antiproliferative effects against tumour cell lines and in animal models of

cancer. AEE788 also exhibited direct antiangiogenic activity. AEE788 is currently in phase I clinical development.

[36] AMN107 is an oral tyrosine kinase inhibitor that targets Bcr-Abl, KIT, and PDGFR. Preclinical studies have shown in cellular assays using Philadelphia chromosome–positive (Ph+) CML cells that AMN107 is highly potent and has high selectivity for Bcr-Abl, KIT, and PDGFR. Weisberg E *et al., Cancer Cell* 7(2):129-41 (February 2005); O'Hare T *et al., Cancer Cell* 7(2):117-9 (February 2005). AMN107 also shows activity against mutated variants of Bcr-Abl. AMN107 is currently being studied in phase I clinical trials.

[37] As used herein, the term "EGFR modulating agent" is any compound that alters (*e.g.*, increases or decreases) the expression level or biological activity level of EGFR polypeptide compared to the expression level or biological activity level of EGFR polypeptide in the absence of the EGFR modulating agent. EGFR modulating agent can be a small molecule, antibody, polypeptide, carbohydrate, lipid, nucleotide, or combination thereof. The EGFR modulating agent can be an organic compound or an inorganic compound.

[38] As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

[39] As used herein, the term "gene" means a segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

[40] As used herein, the term "genotype" means an unphased 5' to 3' sequence of nucleotide pairs found at one or more polymorphic or mutant sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype.

[41] As used herein, the term "locus" means a location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

[42] As used herein, the term "mutant" means any heritable or acquired variation from the wild-type that alters the nucleotide sequence thereby changing the protein sequence. The term "mutant" is used interchangeably with the terms "marker", "biomarker", and "target" throughout the specification.

[43] As used herein, the term "medical condition" includes, but is not limited to, any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment and/or prevention is desirable, and includes previously and newly identified diseases and other disorders.

[44] As used herein, the term "nucleotide pair" means the two nucleotides bound to each other between the two nucleotide strands.

[45] As used herein, the term "polymorphic site" means a position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

[46] As used herein, the term "polymorphism" means any sequence variant present at a frequency of >1% in a population. The sequence variant may be present at a frequency significantly greater than 1% such as 5% or 10 % or more. Also, the term may be used to refer to the sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

[47] As used herein, the term "polynucleotide" means any RNA or DNA, which may be unmodified or modified RNA or DNA. Polynucleotides include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, RNA that is mixture of single- and double-stranded regions, and hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, polynucleotide refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. In a particular embodiment, the polynucleotide contains polynucleotide sequences from the EGFR gene.

[48] As used herein, the term "polypeptide" means any polypeptide comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*, peptide isosteres. Polypeptide refers to both short chains, commonly referred to as peptides, glycopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. Polypeptides include amino acid sequences modified either by natural processes, such as post-

translational processing, or by chemical modification techniques that are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. In a particular embodiment, the polypeptide contains polypeptide sequences from the EGFR protein.

[49] As used herein, the term "small molecule" means a composition that has a molecular weight of less than about 5 kDa and more preferably less than about 2 kDa. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, glycopeptides, peptidomimetics, carbohydrates, lipids, lipopolysaccharides, combinations of these, or other organic or inorganic molecules.

[50] As used herein, the term "mutant nucleic acid" means a nucleic acid sequence, which comprises a nucleotide that is variable within an otherwise identical nucleotide sequence between individuals or groups of individuals, thus, existing as alleles. Such mutant nucleic acids are preferably from about 15 to about 500 nucleotides in length. The mutant nucleic acids may be part of a chromosome, or they may be an exact copy of a part of a chromosome, *e.g.*, by amplification of such a part of a chromosome through PCR or through cloning. The mutant probes according to the invention are oligonucleotides that are complementary to a mutant nucleic acid.

[51] As used herein, the term "SNP nucleic acid" means a nucleic acid sequence, which comprises a nucleotide that is variable within an otherwise identical nucleotide sequence between individuals or groups of individuals, thus, existing as alleles. Such SNP nucleic acids are preferably from about 15 to about 500 nucleotides in length. The SNP nucleic acids may be part of a chromosome, or they may be an exact copy of a part of a chromosome, *e.g.*, by amplification of such a part of a chromosome through PCR or through cloning. The SNP nucleic acids are referred to hereafter simply as "SNPs". The SNP probes according to the invention are oligonucleotides that are complementary to a SNP nucleic acid. In a particular embodiment, the SNP is in the EGFR gene.

[52] As used herein, the term "subject" means that preferably the subject is a mammal, such as a human, but can also be an animal, *e.g.*, domestic animals (*e.g.*, dogs, cats and the like), farm animals (*e.g.*, cows, sheep, pigs, horses and the like) and laboratory animals (*e.g.*, monkey (*e.g.*, cynmologous monkey), rats, mice, guinea pigs and the like).

[53] As used herein, the administration of an agent or drug to a subject or patient includes self-administration and the administration by another. It is also to be appreciated that the

various modes of treatment or prevention of medical conditions as described are intended to mean "substantial", which includes total but also less than total treatment or prevention, and wherein some biologically or medically relevant result is achieved.

[54] The details of one or more embodiments of the invention are set forth in the accompanying description below. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. Other features, objects, and advantages of the invention will be apparent from the description and the claims. In the specification and the appended claims, the singular forms include plural referents unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All references cited herein are incorporated herein by reference in their entireties and for all purposes to the same extent as if each individual publication, patent, or patent application was specifically and individually incorporated by reference in its entirety for all purposes.

[55] *EGFR Mutations and Polymorphisms of the Invention.* A balance between growth-promoting and growth-inhibiting factors regulates cell proliferation and growth. Disturbances in this balance can give rise to uncontrolled cell growth and malignancy, *e.g.*, cancer. A key driver for cell growth is the epidermal growth factor (EGF) and the receptor for EGF ("EGFR"). Goustin *et al.*, *Cancer Res.* 46:1015-1029 (1986); Aaronson SA, *Science* 254:1146-1153 (1991). The EGFR is a transmembrane receptor with an extracellular ligand-binding domain, a helical transmembrane domain, and an intracellular tyrosine kinase domain. Wells A, *Int'l. J. Biochem. Cell Biol.* 31: 637-643 (1999). EGF and other ligands (*e.g.* amphiregulin, TGF-α) bind the EGFR extracellular domain to activate cellular signalling pathways that lead to cell proliferation.

[56] EGFR plays an important role in cellular proliferation as well as apoptosis, angiogenesis and metastatic spread, processes that are crucial to tumour progression. Salomon *et al.*, *Crit. Rev. Oncology/Haematology*, 19:183-232 (1995); Wu *et al.*, *J. Clin. Invest.*, 95:1897-1905 (1995); Karnes *et al.*, *Gastroenterology*, 114:930-939 (1998); Woodburn *et al.*, *Pharmacol. Therap.* 82: 241-250 (1999); Price *et al.*, *Eur. J. Cancer*, 32A:1977-1982 (1996). Indeed, studies have shown that EGFR-mediated cell growth is increased in a variety of solid tumours including non-small cell lung cancer, prostate cancer, breast cancer, gastric cancer,

and tumours of the head and neck. Salomon DS *et al.*, *Critical Reviews in Oncology/Haematology*, 19:183-232 (1995). Furthermore, excessive activation of EGFR on the cancer cell surface is now known to be associated with advanced disease, the development of a metastatic phenotype and a poor prognosis in cancer patients. Salomon DS *et al.*, *Critical Reviews in Oncology/Haematology* 19:183-232 (1995).

[57] Mutations of EGFR have been identified in human cancer patients that affect their response to chemotherapy directed toward EGFR. For example, studies by Lynch and co-workers (*N. Engl. J. Med.* 350: 2129-2139 (2004)) and Paez and co-workers (*Science* 304: 1497-1500 (2004)) have described somatic mutations in the EGFR gene in patients with non-small cell lung cancer (NSCLC) who were particularly responsive to gefitinib, an EGFR kinase inhibitor. Although different types of mutations were identified, they all clustered around the ATP-binding pocket of the receptor's tyrosine kinase domain and have been shown to enhance sensitivity to gefitinib in preclinical models. EGFR mutations have also been identified in patients experiencing responses to the EGFR tyrosine kinase inhibitor erlotinib. Shepherd *et al.*, *Proc. Am. Soc. Clin. Oncol.*, 23: 18 (abstr. # 7022) (2004).

[58] *Identification of EGFR Mutations and Polymorphisms of the Invention in Human Cancers.* To determine the EGFR mutations and polymorphisms in a variety of human tumours, DHPLC analysis (Lilleberg SL, *Curr. Opin. Drug Discov. Devel.* 6(2): 237-52 (March 2003)) was conducted on tissue samples and cell lines derived from human cancers, *e.g.*, glioblastoma, breast cancer, cholangioma, non-small-cell lung cancer (NSCLC), prostate cancer, colon cancer, medullary thyroid cancer, melanoma, ovarian cancer and myeloma as summarized in TABLE 1 below. As shown in TABLE 1, forty-three (43) missense mutations were identified in the EGFR gene (NT 079592). Some missense mutations are identified in two or more types of tumour types, *e.g.* P699A in medullary thyroid cancer and melanoma, V738G in ovarian cancer and NSCLC, K754R in cholangioma, breast cancer and prostate cancer, K757R in cholangioma and prostate cancer, G779S in cholangioma and prostate cancer, T751I in prostate cancer and myeloma. None of these mutations has been reported previously.

TABLE 1
Missense Mutations of EGFR

	Thyroid cancer	Melanoma	Colorectal cancer	Glioblastoma	Cholangioma	Ovarian Cancer	Breast cancer	Prostate Cancer	NSCLC	Myeloma
Q217R			X							
G221W							X			
G239C						X				
C251F							X			
ins R255-T290			X							
L267V					X					
E282D						X				
D587N						X				
L509V						X				
C595W							X			
H618Y								X		
T693I									X	
P699A	X	X								
E709V	X									
K716N								X		
T725P						X				
Y727S								X		
G729R								X		
I732T									X	
G735R				X						
E736D				X						
V738G						X			X	
I740T									X	
K745R					X					
E746G									X	
R748G					X					
R748I							X			
T751I								X		X
T751A									X	
K754E							X			
K754R					X		X	X		
N756S									X	
K757R				X		X				
K757E								X		
E758G							X			
L760P				X						
N771D				X						
V774M									X	
G779S					X		X			

TABLE 1
Missense Mutations of EGFR

	Thyroid cancer	Melanoma	Colorectal cancer	Glioblastoma	Cholangioma	Ovarian Cancer	Breast cancer	Prostate Cancer	NSCLC	Myeloma
I890T				X						
Y891S			X							
Q894X					X					
Y900X						X				
R932G									X	
M971R							X			

[59] Apart from mutations, eighteen (18) SNPs were identified in the EGFR gene as shown in TABLE 2. Ten (10) SNPs at T263A, Y610Y, A613A, L683L, K737K, P741P, L747L, K754K, T785T, and D916D have not been reported previously.

TABLE 2
SNPs Identified In the Coding Region of Human EGFR
Numbers Represent Frequency of SNPs Observed

	Colorectal	Glioblastoma	MTC	Melanoma	NSCLC	Ovarian	Prostate	Breast	Cholangioma
N158N	0.53	0.40	0.47	0.53	0.53	0.47	0.47	0.48	0.47
T263A						0.07	0.03		0.03
P373P	0.17	0.23	0.13	0.20	0.20	0.13	0.17	0.16	0.13
R521K	0.40	0.37	0.40	0.40	0.50	0.13	0.30	0.39	0.33
Y610Y				0.13	0.13			0.03	
A613A	0.13	0.43	0.47	0.27	0.23	0.13	0.17	0.03	0.17
T629T	0.67	0.57	0.67	0.73	0.70	0.27	0.27	0.03	0.50
L683L			0.07		0.10				
K737K							0.03		
P741P		0.03							
L747L								0.03	
K754K						0.13	0.03		
Q787Q	0.47	0.67	0.47	0.40	0.73	0.27	0.47	0.03	0.43
T785T		0.03							
R836R	0.03								
T903T		0.13	0.13	0.13	0.13	0.27	0.17	0.03	0.13
D916D					0.03				
D994D	0.17	0.13	0.13	0.13	0.13	0.10	0.10		0.10

[60] TABLE 3 summarizes sequence variations that have been identified in this invention.

TABLE 3
EGFR Mutations and Polymorphisms in Selected Cancers

	NT change	AA change	Allelic frequency	Number per 30 tumours
Breast	AAC>AAT	N158N	het	15
	GGG>TGG	G221W	<0.1	1
	TGC>TTC	C251F	0.2	1
	AGG>AAG	R521K	het	1
	TGC>TGG	C595W	<0.1	1
	AAA>GAA	K754E	0.15	1
	AGA>ATA	R748I	<0.1	1
	AAA>AGA	K754R	0.1	1
	ATG>AGG	M952R	0.2	1
	GCC>GCT	A613A	het	1
	ACT>ACA	T629T	het	15
	TTA>CTA	L747L	het	1
	CAA>CAG	Q787Q	het	9
	G>A	P373P	het	1
	CTG>TTG	L683L	het	2
	GAA>GAG	E734E	0.2	1
	GAA>GAG	E758G	0.2	1
	GGC>AGC	G779S	0.4	1
	ACC>ACT	T903T	het	10
Ovarian	GGC>TGC	G239C	<0.1	1
	GAG>GAT	E282D	0.15	1
	ACC>GCC	T263A	het	2
	AGG>AAG	R521K	het	4
	TTG>GTG	L509V	<0.1	1
	GCC>GCT	A613A	het	4
	GAC>AAC	D587N	0.3	1
	ACT>ACA	T629T	het	8
	ACG>CCG	T725P	0.35	3
	AAC>AAT	N158N	het	14
	AAA>AAG	K739K	het	1
	AAG>AGA	K757R	0.25	1
	GTT>GGT	V738G	0.1	1
	AAA>AAG	K754K	het	3
	CAA>CAG	Q787Q	het	8
	TAT>TAG	Y900X	0.1	1
	ACC>ACT	T903T	het	8
	C>T	D994D	het	4
	ACC>ACT	T903T	het	8
	ACC>ACT	T903T	het	8

TABLE 3
EGFR Mutations and Polymorphisms in Selected Cancers

	NT change	AA change	Allelic frequency	Number per 30 tumours
Thyroid	AAC>AAT	N158N	het	7
	GCC>GCT	A613A	het	3
	CTG>TTG	L683L	het	1
	CCC>GCC	P699A	<0.1	1
	CCA>CAG	Q787Q	het	7
	G>A	P373P	het	2
	AGG>AAG	R521K	hom, het	6
	ACT>ACA	T629T	hom, het	10
	CTG>TTG	L683L	het	1
	GAA>GTA	E709V	0.12	1
	C>T	D994D	het	2
	ACC>ACT	T903T	het	2
Colorectal	AAC>AAT	N158N	het	16
	CAG>CGG	Q217R	0.2	1
	ins 105bp	R255-T290	0.15	1
	AGG>AAG	R521K	het	12
	GCC>GCT	A613A	het	14
	ACT>ACA	T629T	het	12
	CAA>CAG	Q787Q	het	14
	CGC>CGT	R836R	het	1
	G>A	P373P	het	5
	C>T	D994D	het	5
	TAT>TCT	Y891S	<0.1	1
Melanoma (15)	AAC>AAT	N158N	het	8
	GCC>GCT	A613A	het	4
	CCC>GCC	P699A	0.05	1
	CAA>CAG	Q787Q	het	6
	G>A	P373P	het	3
	AGG>AAG	R521K	het	6
	TAC>TAT	Y610Y	het	2
	ACT>ACA	T629T	hom, het	11
	C>T	D994D	het	2
	ACC>ACT	T903T	het	2

TABLE 3

EGFR Mutations and Polymorphisms in Selected Cancers

	NT change	AA change	Allelic frequency	Number per 30 tumours
Prostate	ACC>GCC	T263A	het	1
	AGG>AAG	R521K	het	9
	GCC>GCT	A613A	het	4
	CAC>TAC	H618Y	0.1	1
	ACT>ACA	T629T	het	8
	GCT>GCG	A698A	0.1	1
	ATC>ACC	I706T	0.6	1
	AAA>AAC	K716N	0.1	1
	TAT>TCT	Y727S	0.4	1
	GGA>AGA	G729R	<0.1	1
	AAA>AAG	K737K	het	1
	ACA>ATA	T751I	0.15	1
	AAA>AAG	K754K	het	1
	AAA>AGA	K754R	0.1	1
	T/C	2 bp 3' from SD	0.25	1
	CAA>CAG	Q787Q	het	14
	ACC>ACT	T903T	het	5
	GAC>GAT	D909D	het	1
	AAC>AAT	N158N	het	14
	G>A	P373P	het	5
	AAG>GAG	K757E	<0.1	1
	C>T	D994D	het	3
Glioblastoma	AAC>AAT	N158N	het	13
	G>A	P373P	het	8
	AGG>AAG	R521K	het	12
	GCC>GCT	A613A	hom, het	14
	ACT>ACA	T619T	het	17
	GGT>CGT	G735R	0.1	1
	CTC>CCC	L760P	0.25	1
	CCC>CCT	P741P	het	1
	T>A	13bp 3' SD	het	1
	GAG>GAC	E736D	0.1	1
	AAG>AGA	K757R	0.25	1
	CAA>CAG	Q787Q	het	20
	AAC>GAC	N771D	0.07	1
	ACC>ACT	T785T	het	1
	ATC>ACC	I890T	0.2	1
	ACC>ACT	T903T	het	5
	C>T	D994D	het	4

TABLE 3
EGFR Mutations and Polymorphisms in Selected Cancers

	NT change	AA change	Allelic frequency	Number per 30 tumours
Cholangioma	AAC>AAT	N158N	het	14
	ACC>GCC	T263A	het	1
	CTC>GTC	L267V	0.15	1
	G>A	P373P	het	4
	AGG>AAG	R521K	het	10
	GCC>GCT	A613A	het	5
	ACT>ACA	T629T	hom, het	15
	AGA>GGA	R748G	0.2	1
	AAA>AGA	K754R	0.1	1
	AAG>AGG	K45R	0.3	1
	GAT>GAC	D761D	0.2	1
	CAA>CAG	Q787Q	het	13
	GGC>AGC	G779S	0.2	1
	CAG>TAG	Q894Term	0.4	1
	ACC>ACT	T903T	het	4
	C>T	D994D	het	3
NSCLC	AAC>AAT	N158N	het	16
	G>A	P373P	het	6
	AGG>AAG	R521K	het	15
	TAC>TAT	Y610Y	het	4
	GCC>GCT	A613A	het	7
	ACT>ACA	T629T	het	21
	A>G	2bp 5' to SD	0.4	1
	C>T	L683L	het	3
	ATT>ACT	I732T	0.1	1
	GAA>GGA	E746G	0.6	1
	ACA>GCA	T751A	0.1	1
	ACA>ATA	T693I	0.15	1
	ATT>ACT	I740T	0.1	1
	GTT>GGT	V738G	0.3	1
	AAC>AGC	N756S	0.15	1
	CAA>CAG	Q787Q	het	11
	GTG>ATG	V774M	0.1	1
	ACC>ACT	T903T	het	4
	CGC>GGC	R732G	0.15	1
	GAC>GAT	D916D	het	1
	C>T	D994D	het	4
Myeloma (cell line RPMI8226)	ACA>ATA	T751I	het	1

[61] The human EGFR wild-type polypeptide sequence (SEQ ID NO:1) is shown below in TABLE 4. The amino acid residues encoded at the position of the EGFR missense mutations of the EGFR gene summarized in TABLE 3 are highlighted in bold underlined text below in TABLE 4.

TABLE 4

Human EGFR (isoform a) Wild-type Polypeptide Sequence

```
   1 mrpsgtagaa llallaalcp asraleekkv cqgtsnkltq lgtfedhfls lqrmfnncev
  61 vlgnleityv qrnydlsflk tiqevagyvl ialntverip lenlqiirgn myyensyala
 121 vlsnydankt glkelpmrnl qeilhgavrf snnpalcnve siqwrdivss dflsnmsmdf
 181 qnhlgscqkc dpscpngscw gageencqkl tkiicaqqcs grcrgkspsd cchnqcaagc
 241 tgpresdclv crkfrdeatc kdtcpplmly npttyqmdvn pegkysfgat cvkkcprnyv
 301 vtdhgscvra cgadsyemee dgvrkckkce gpcrkvcngi gigefkdsls inatnikhfk
 361 nctsisgdlh ilpvafrgds fthtppldpq eldilktvke itgflliqaw penrtdlhaf
 421 enleiirgrt kqhgqfslav vslnitslgl rslkeisdgd viisgnknlc yantinwkkl
 481 fgtsgqktki isnrgensck atgqvchalc spegcwgpep rdcvscrnvs rgrecvdkcn
 541 llegeprefv enseciqchp eclpqamnit ctgrgpdnci qcahyidgph cvktcpagvm
 601 genntlvwky adaghvchlc hpnctygctg pglegcptng pkipsiatgm vgallllvv
 661 algiglfmrr rhivrkrtlr rllqerelve pltpsgeapn qallrilket efkkikvlgs
 721 gafgtvykgl wipegekvki pvaikelrea tspkankeil deayvmasvd nphvcrllgi
 781 cltstvqlit qlmpfgclld yvrehkdnig sqyllnwcvq iakgmnyled rrlvhrdlaa
 841 rnvlvktpqh vkitdfglak llgaeekeyh aeggkvpikw malesilhri ythqsdvwsy
 901 gvtvwelmtf gskpydgipa seissilekg erlpqppict idvymimvkc wmidadsrpk
 961 freliiefsk mardpqrylv iqgdermhlp sptdsnfyra lmdeedmddv vdadeylipq
1021 qgffsspsts rtpllsslsa tsnnstvaci drnglqscpi kedsflqrys sdptgalted
1081 siddtflpvp eyinqsvpkr pagsvqnpvy hnqplnpaps rdphyqdphs tavgnpeyln
1141 tvqptcvnst fdspahwaqk gshqisldnp dyqqdffpke akpngifkgs taenaeylrv
1201 apqssefiga (SEQ ID NO:1)
```

[62] Studies to determine EGFR mutations in colorectal cancer and lung cancer are summarized in EXAMPLE 1. Bioinformatics analyses of the EGFR mutations of the invention are further detailed in EXAMPLE 2.

[63] *Identification and Characterization of Gene Sequence Variation.* Sequence variation in the human germline consists primarily of SNPs, the remainder being short tandem repeats (including micro-satellites), long tandem repeats (mini-satellites), and other insertions and deletions. A SNP is the occurrence of nucleotide variability at a single position in the genome, in which two alternative bases occur at appreciable frequency (*i.e.*, >1%) in the human population. A SNP may occur within a gene or within intergenic regions of the genome.

[64] Due to their prevalence and widespread nature, SNPs have the potential to be important tools for locating genes that are involved in human disease conditions. See *e.g.*, Wang *et al., Science* 280: 1077-1082 (1998)).

[65] An association between SNP's and/or mutations and a particular phenotype (*e.g.* cancer type) does not necessarily indicate or require that the SNP or mutation is causative of the phenotype. Instead, an association with a SNP may merely be due to genome proximity between a SNP and those genetic factors actually responsible for a given phenotype, such that the SNP and said genetic factors are closely linked. That is, a SNP may be in linkage disequilibrium ("LD") with the "true" functional variant. LD exists when alleles at two distinct locations of the genome are more highly associated than expected. Thus, a SNP may serve as a marker that has value by virtue of its proximity to a mutation or other DNA alteration (e.g. gene duplication) that causes a particular phenotype.

[66] SNPs and mutations that are associated with disorders may also have a direct effect on the function of the genes in which they are located. For example, a sequence variant (e.g. SNP) may result in an amino acid change or may alter exon-intron splicing, thereby directly modifying the relevant protein, or it may exist in a regulatory region, altering the cycle of expression or the stability of the mRNA (*see*, *e.g.*, Nowotny *et al.*, *Current Opinions in Neurobiology*, 11:637-641 (2001)).

[67] In describing the polymorphic and mutant sites of the invention, reference is made to the sense strand of the gene for convenience. As recognized by the skilled artisan, however, nucleic acid molecules containing the gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. That is, reference may be made to the same polymorphic or mutant site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic and/or mutant site. Thus, the invention also includes single-stranded polynucleotides and mutations that are complementary to the sense strand of the genomic variants described herein.

[68] *Identification and Characterization of SNPs and Mutations* .Many different techniques can be used to identify and characterize SNPs and mutations, including single-strand conformation polymorphism (SSCP) analysis, heteroduplex analysis by denaturing high-performance liquid chromatography (DHPLC), direct DNA sequencing and computational methods (Shi *et al.*, *Clin Chem* 47:164-172 (2001)). There is a wealth of sequence information in public databases; computational tools useful to identify SNPs *in silico* by aligning independently submitted sequences for a given gene (either cDNA or genomic sequences) The most common SNP-typing methods currently include hybridization, primer

extension, and cleavage methods. Each of these methods must be connected to an appropriate detection system. Detection technologies include fluorescent polarization (Chan *et al., Genome Res.* 9:492-499 (1999)), luminometric detection of pyrophosphate release (pyrosequencing) (Ahmadiian *et al., Anal. Biochem.* 280:103-10 (2000)), fluorescence resonance energy transfer (FRET)-based cleavage assays, DHPLC, and mass spectrometry (Shi, *Clin Chem* 47:164-172 (2001); U.S. Pat. No. 6,300,076 B1). Other methods of detecting and characterizing SNPs and mutations are those disclosed in U.S. Pat. Nos. 6,297,018 B1 and 6,300,063 B1.

[69] In a particularly preferred embodiment, the detection of polymorphisms and mutations is detected using INVADER™ technology (available from Third Wave Technologies Inc. Madison, Wisconsin USA). In this assay, a specific upstream "invader" oligonucleotide and a partially overlapping downstream probe together form a specific structure when bound to complementary DNA template. This structure is recognized and cut at a specific site by the Cleavase enzyme, resulting in the release of the 5' flap of the probe oligonucleotide. This fragment then serves as the "invader" oligonucleotide with respect to synthetic secondary targets and secondary fluorescently labelled signal probes contained in the reaction mixture. This results in specific cleavage of the secondary signal probes by the Cleavase enzyme. Fluorescent signal is generated when this secondary probe (labelled with dye molecules capable of fluorescence resonance energy transfer) is cleaved. Cleavases have stringent requirements relative to the structure formed by the overlapping DNA sequences or flaps and can, therefore, be used to specifically detect single base pair mismatches immediately upstream of the cleavage site on the downstream DNA strand. Ryan D *et al., Molecular Diagnosis* 4(2): 135-144 (1999) and Lyamichev V *et al.. Nature Biotechnology* 17: 292-296 (1999), see also U.S. Pat. Nos. 5,846,717 and 6,001,567.

[70] The identity of polymorphisms and mutations may also be determined using a mismatch detection technique including, but not limited to, the RNase protection method using riboprobes (Winter *et al., Proc. Natl. Acad. Sci. USA* 82:7575 (1985); Meyers *et al., Science* 230:1242 (1985)) and proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich P, *Ann Rev Genet* 25:229-253 (1991)). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita *et al., Genomics* 5:874-879 (1989); Humphries *et al.*, in *Molecular Diagnosis of Genetic Diseases*, Elles R, ed. (1996) pp. 321-340) or denaturing gradient gel electrophoresis (DGGE)

(Wartell *et al., Nucl. Acids Res.* 18:2699-2706 (1990); Sheffield *et al., Proc. Natl. Acad. Sci. USA* 86: 232-236 (1989)). A polymerase-mediated primer extension method may also be used to identify the polymorphisms/mutations. Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO 92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Pat. No. 5,679,524). Related methods are disclosed in WO 91/02087, WO 90/09455, WO 95/17676, and U.S. Pat. Nos. 5,302,509 and 5,945,283. Extended primers containing a polymorphism or mutation may be detected by mass spectrometry as described in U.S. Pat. No. 5,605,798. Another primer extension method is allele-specific PCR. Ruafio *et al., Nucl. Acids Res.* 17: 8392 (1989); Ruafio *et al., Nucl. Acids Res.* 19: 6877-6882 (1991); WO 93/22456; Turki *et al., J. Clin. Invest.* 95: 1635-1641 (1995). In addition, multiple polymorphic and/or mutant sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in WO 89/10414.

[71] *Haplotyping and Genotyping Oligonucleotides.* The invention provides methods and compositions for haplotyping and/or genotyping the genetic polymorphisms (and possibly mutations) in an individual. As used herein, the terms "genotype" and "haplotype" mean the genotype or haplotype containing the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic (or mutant) sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic (or mutant) sites in the gene. The additional polymorphic (and mutant) sites may be currently known polymorphic/mutant sites or sites that are subsequently discovered.

[72] The compositions contain oligonucleotide probes and primers designed to specifically hybridize to one or more target regions containing, or that are adjacent to, a polymorphic or mutant site. Oligonucleotide compositions of the invention are useful in methods for genotyping and/or haplotyping a gene in an individual. The methods and compositions for establishing the genotype or haplotype of an individual at the novel polymorphic/mutant sites described herein are useful for studying the effect of the polymorphisms and mutations in the aetiology of diseases affected by the expression and function of the protein, studying the efficacy of drugs targeting, predicting individual susceptibility to diseases affected by the expression and function of the protein and predicting individual responsiveness to drugs targeting the gene product.

[73] Some embodiments of the invention contain two or more differently labelled genotyping oligonucleotides, for simultaneously probing the identity of nucleotides at two or more polymorphic or mutant sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic or mutant site.

[74] Genotyping oligonucleotides of the invention may be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, *e.g.*, WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism and mutation detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms and mutations in multiple genes at the same time.

[75] An allele-specific oligonucleotide primer of the invention has a 3' terminal nucleotide, or preferably a 3' penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. Allele-specific oligonucleotide (ASO) primers hybridizing to either the coding or noncoding strand are contemplated by the invention. An ASO primer for detecting gene polymorphisms and mutations can be developed using techniques known to those of skill in the art.

[76] Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic or mutant sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms or mutations described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic/mutant site.

[77] *Direct Genotyping Method of the Invention.* One embodiment of a genotyping method of the invention involves isolating from an individual a nucleic acid mixture comprising at least one copy of the gene of interest and/or a fragment or flanking regions thereof, and determining the identity of the nucleotide pair at one or more of the polymorphic/mutant sites in the nucleic acid mixture. As will be readily understood by the skilled artisan, the two

"copies" of a germline gene in an individual may be the same on each allele or may be different on each allele. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each polymorphic and mutant site.

[78] Typically, the nucleic acid mixture is isolated from a biological sample taken from the individual, such as a blood sample, tumour or tissue sample. Suitable tissue samples include whole blood, tumour or as part of any tissue type, semen, saliva, tears, urine, fecal material, sweat, buccal smears, skin and hair. The nucleic acid mixture may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from an organ in which the gene may be expressed. Furthermore, it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms or mutations located in introns or in 5' and 3' nontranscribed regions. If a gene fragment is isolated, it must usually contain the polymorphic and/or mutant sites to be genotyped. Exceptions can include mutations leading to truncation of the gene where a specific polymorphism may be lost. In these cases, the specific DNA alterations are determined by assessing the flanking sequences of the gene and underscore the need to specifically look for both polymorphisms and mutations.

[79] *Direct Haplotyping Method of the Invention.* One embodiment of the haplotyping method of the invention comprises isolating from an individual a nucleic acid molecule containing only one of the two copies of a gene of interest, or a fragment thereof, and determining the identity of the nucleotide at one or more of the polymorphic or mutant sites in that copy. The nucleic acid may be isolated using any method capable of separating the two copies of the gene or fragment. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the gene in an individual. In a particularly preferred embodiment, the nucleotide at each polymorphic or mutant site is identified.

[80] In a preferred embodiment, a haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more of the polymorphic/mutant sites in each copy of the gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of

nucleotides at each polymorphic/mutant site in each copy of the gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, if the two copies are labelled with different tags, or are otherwise separately distinguishable or identifiable, it is possible in some cases to perform the method in the same container. For example, if the first and second copies of the gene are labelled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labelled with yet a third different fluorescent dye is used to assay the polymorphic/mutant sites, then detecting a combination of the first and third dyes would identify the polymorphism or mutation in the first gene copy, while detecting a combination of the second and third dyes would identify the polymorphism or mutation in the second gene copy.

[81] In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic and/or mutant site may be determined by amplifying a target region containing the polymorphic and/or mutant sites directly from one or both copies of the gene, or fragments thereof, and sequencing the amplified regions by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic or mutant site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism or mutation may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for all individuals homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

[82] *Indirect Genotyping Method using Polymorphic and Mutation Sites in Linkage Disequilibrium with a Target Polymorphism or Mutation.* In addition, the identity of the alleles present at any of the novel polymorphic/mutant sites of the invention may be indirectly determined by genotyping other polymorphic/mutant sites in linkage disequilibrium with those sites of interest. As described *supra*, two sites are said to be in linkage disequilibrium if the presence of a particular variant (polymorphism or mutation) at one site is indicative of the presence of another variant at a second site. *See*, Stevens JC, *Mol. Diag.* 4:309-317 (1999).

Polymorphic and mutant sites in linkage disequilibrium with the polymorphic or mutant sites of the invention may be located in regions of the same gene or in other genomic regions. Genotyping of a polymorphic/mutant site in linkage disequilibrium with the novel polymorphic/mutant sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic/mutant site.

[83] *Amplifying a Target Gene Region.* The target regions may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR). (U.S. Pat. No. 4,965,188), ligase chain reaction (LCR) (Barany *et al., Proc. Natl. Acad. Sci. USA* 88:189-193 (1991); published PCT patent application WO 90/01069), and oligonucleotide ligation assay (OLA) (Landegren *et al., Science* 241: 1077-1080 (1988)). Oligonucleotides useful as primers or probes in such methods should specifically hybridize to a region of the nucleic acid that contains or is adjacent to the polymorphic/mutant site. Typically, the oligonucleotides are between 10 and 35 nucleotides in length and preferably, between 15 and 30 nucleotides in length. Most preferably, the oligonucleotides are 20 to 25 nucleotides long. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan.

[84] Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Pat. No. 5,130,238; EP 329,822; U.S. Pat. No. 5,169,766, published PCT patent application WO 89/06700) and isothermal methods (Walker *et al., Proc. Natl. Acad. Sci. USA* 89: 392-396 (1992)).

[85] A polymorphism or mutation in the target region may be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labelled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic/mutant site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic or mutant sites being detected.

[86] *Hybridizing Allele-Specific Oligonucleotide to a Target Gene.* Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking, baking, *etc*. Allele-specific oligonucleotide may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibres, chips, dishes, and beads. The solid support may be treated, coated or derivatised to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

[87] The genotype or haplotype for the gene of an individual may also be determined by hybridization of a nucleic sample containing one or both copies of the gene to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic or mutant sites to be included in the genotype or haplotype.

[88] *Determining Population Genotypes and Haplotypes and Correlating them with a Trait.* The present invention provides a method for determining the frequency of a genotype or haplotype in a population. The method comprises determining the genotype or the haplotype for a gene present in each member of the population, wherein the genotype or haplotype comprises the nucleotide pair or nucleotide detected at one or more of the polymorphic sites in the gene and mutations identified in the region, and calculating the frequency at which the genotype or haplotype is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (*e.g.*, a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

[89] In another aspect of the invention, frequency data for genotypes and/or haplotypes found in a reference population are used in a method for identifying an association between a trait and a genotype or a haplotype. The trait may be any detectable phenotype, including but not limited to cancer, susceptibility to a disease or response to a treatment. The method

involves obtaining data on the frequency of the genotypes or haplotypes of interest in a reference population and comparing the data to the frequency of the genotypes or haplotypes in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above.

[90] In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. Such methods have applicability in developing diagnostic tests and therapeutic treatments for all pharmacogenetic applications where there is the potential for an association between a genotype and a treatment outcome, including efficacy measurements, PD measurements, PK measurements and side effect measurements.

[91] In another embodiment, the frequency data for the reference and/or trait populations are obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data are obtained, the frequencies of the genotypes or haplotypes of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes and/or haplotypes observed in the populations are compared. If a particular genotype or haplotype for the gene is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that genotype or haplotype.

[92] In a preferred embodiment, the haplotype frequency data for different ethnogeographic groups are examined to determine whether they are consistent with Hardy-Weinberg equilibrium. Hartl DL *et al., Principles of Population Genomics, 3rd Ed.* (Sinauer Associates, Sunderland, MA, 1997). Hardy-Weinberg equilibrium postulates that the frequency of finding the haplotype pair H_1/H_2 is equal to $P_{H\text{-}W}(H_1/H_2) = 2p(H_1)\,p(H_2)$ if $H_1 \neq H_2$ and $P_{H\text{-}W}(H_1/H_2) = p(H_1)\,p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not

reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Pat. No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin *et al., Nucl. Acids Res.* 24: 4841-4843 (1996)).

[93] In one embodiment of this method for predicting a haplotype pair, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, those discussed *supra*.

[94] In a preferred embodiment, statistical analysis is performed by the use of standard ANOVA tests with a Bonferoni correction and/or a bootstrapping method that simulates the genotype phenotype correlation many times and calculates a significance value. When many polymorphisms and/or mutations are being analyzed, a calculation may be performed to correct for a significant association that might be found by chance. For statistical methods useful in the methods of the present invention, see Bailey NTJ, *Statistical Methods in Biology, 3rd Edition* (Cambridge Univ. Press, Cambridge, 1997); Waterman MS, *Introduction to Computational Biology* (CRC Press, 2000) and *Bioinformatics*, Baxevanis AD & Ouellette BFF, eds. (John Wiley & Sons, Inc., 2001).

[95] In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting or to a therapeutic treatment for a medical condition.

[96] In another embodiment of the invention, a detectable genotype or haplotype that is in linkage disequilibrium with a genotype or haplotype of interest may be used as a surrogate marker. A genotype that is in linkage disequilibrium with another genotype is indicated where a particular genotype or haplotype for a given gene is more frequent in the population

that also demonstrates the potential surrogate marker genotype than in the reference population. If the frequency is statistically significant, then the marker genotype is predictive of that genotype or haplotype, and can be used as a surrogate marker.

[97] *Correlating Subject Genotype or Haplotype to Treatment Response.* In order to deduce a correlation between a clinical response to a treatment and a genotype or haplotype, genotype or haplotype data is obtained on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been previously conducted and/or by designing and carrying out one or more new clinical trials.

[98] It is preferred that the individuals included in the clinical population be graded for the existence of the medical condition of interest. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use genotyping or haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

[99] The therapeutic treatment of interest is administered to each individual in the trial population, and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses, and that the investigator may choose more than one responder groups (*e.g.*, low, medium, high) made up by the various responses. In addition, the gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

[100] These results are then analyzed to determine if any observed variation in clinical response between polymorphism/mutation groups is statistically significant. Statistical analysis methods, which may be used, are described in Fisher LD & vanBelle G, *Biostatistics: A Methodology for the Health Sciences* (Wiley-Interscience, New York, 1993). This analysis may also include a regression calculation of which polymorphic/mutation sites in the gene contribute most significantly to the differences in phenotype.

[101] A second method for finding correlations between genotype and haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms, one of which is a genetic algorithm (Judson R, *Genetic Algorithms and Their Uses in Chemistry*, in *Reviews in Computational Chemistry, Vol. 10*, Lipkowitz KB & Boyd DB, eds. (VCH Publishers, New York, 1997) pp. 1- 73. Simulated annealing (Press *et al.*, *Numerical*

Recipes in C: The Art of Scientific Computing, Ch. 10 (Cambridge University Press, Cambridge, 1992)), neural networks (Rich E & Knight K, *Artificial Intelligence, 2nd Edition, Ch. 10* (McGraw-Hill, New York, 1991), standard gradient descent methods (Press *et al.*, *Numerical Recipes in C: The Art of Scientific Computing, Ch. 10* (Cambridge University Press, Cambridge, 1992), or other global or local optimization approaches (see discussion in Judson, supra) can also be used.

[102] Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic and mutant sites in the gene. ANOVA is used to test hypotheses about whether a response variable is caused by or correlates with one or more traits or variables that can be measured (Fisher & vanBelle, *supra,* Ch. 10).

[103] After the clinical, mutation and polymorphism data have been obtained, correlations between individual response and genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their genotype or haplotype (or haplotype pair) (also referred to as a polymorphism/mutation group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism/mutation group are calculated.

[104] From the analyses described above, the skilled artisan that predicts clinical response as a function of genotype or haplotype content may readily construct a mathematical model. The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, *i.e.*, a greater dose of a drug or suffer an adverse reaction. The diagnostic method may take one of several forms: for example, a direct DNA test (*i.e.*, genotyping or haplotyping one or more of the polymorphic/mutant sites in the gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying genotype or haplotype. In a preferred embodiment, this diagnostic method uses the predictive genotyping/haplotyping method described above.

[105] *Patient Selection for Therapy Based Upon Polymorphisms and/or Mutations.* The application of genotypes and/or haplotypes that correlate with efficacious drug responses will be used to select patients for therapy of existing diseases. Genotypes and haplotypes that

correlate with adverse consequences will be used to either modify how the drug is administered (e.g. dose, schedule or in combination with other drugs) or eliminated as an option.

[106] *Patient Selection for Prophylactic Therapy Based Upon Polymorphisms and/or Mutations.* The application of genotypes and/or haplotypes that correlate with a predisposition for disease will be used to select patients for preventative therapy..

[107] *Computer System for Storing or Displaying Polymorphism and Mutation Data.* The invention also provides a computer system for storing and displaying polymorphism and mutation data determined for the gene. The computer system comprises a computer processing unit, a display, and a database containing the polymorphism/mutation data. The polymorphism/mutation data includes the polymorphisms, mutations, the genotypes and the haplotypes identified for a given gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing haplotypes organized according to their evolutionary relationships. A computer may implement any or all analytical and mathematical operations involved in practicing the methods of the present invention. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, mutation data, genetic sequence data, and clinical population data (*e.g.*, data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations).. The polymorphism and mutation data described herein may be stored as part of a relational database (*e.g.*, an instance of an Oracle database or a set of ASCII flat files). These polymorphism and mutation data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer *via* a network.

[108] *Nucleic Acid-based Diagnostics.* In another aspect, the invention provides SNP and mutation probes, which are useful in classifying subjects according to their types of genetic variation. The SNP and mutation probes according to the invention are oligonucleotides, which discriminate between SNPs or mutations and the wild-type sequence in conventional allelic discrimination assays. In certain preferred embodiments, the oligonucleotides

according to this aspect of the invention are complementary to one allele of the SNP/mutant nucleic acid, but not to any other allele of the SNP/Mutant nucleic acid. Oligonucleotides according to this embodiment of the invention can discriminate between SNPs and mutations in various ways. For example, under stringent hybridization conditions, an oligonucleotide of appropriate length will hybridize to one SNP or mutation, but not to any other. The oligonucleotide may be labelled using a radiolabel or a fluorescent molecular tag. Alternatively, an oligonucleotide of appropriate length can be used as a primer for PCR, wherein the 3' terminal nucleotide is complementary to one allele containing a SNP or mutation, but not to any other allele. In this embodiment, the presence or absence of amplification by PCR determines the haplotype of the SNP or the specific mutation.

[109] Genomic and cDNA fragments of the invention comprise at least one novel polymorphic site or mutation identified herein, have a length of at least 10 nucleotides, and may range up to the full length of the gene. Preferably, a fragment according to the present invention is between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

[110] *Kits of the Invention.* The invention provides nucleic acid and polypeptide detection kits useful for haplotyping and/or genotyping the genes in an individual. Such kits are useful for classifying individuals for the purpose of classifying individuals. Specifically, the invention encompasses kits for detecting the presence of a polypeptide or nucleic acid corresponding to a marker of the invention in a biological sample, *e.g.*, any tissue or bodily fluid including, but not limited to, serum, plasma, lymph, cystic fluid, urine, stool, cerebrospinal fluid, ascites fluid or blood, and including biopsy samples of body tissue. For example, the kit can comprise a labelled compound or agent capable of detecting a polypeptide or an mRNA encoding a polypeptide corresponding to a marker of the invention in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample, *e.g.*, an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide. Kits can also include instructions for interpreting the results obtained using the kit.

[111] In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a

probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as in the case of PCR.

[112] In a preferred embodiment, such kit may further comprise a DNA sample collecting means. In particular, the genotyping primer composition may comprise at least two sets of allele specific primer pairs. Preferably, the two genotyping oligonucleotides are packaged in separate containers.

[113] For antibody-based kits, the kit can comprise, *e.g.*, (1) a first antibody, *e.g.*, attached to a solid support, which binds to a polypeptide corresponding to a marker or the invention; and, optionally (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable label.

[114] For oligonucleotide-based kits, the kit can comprise, *e.g.*, (1) an oligonucleotide, *e.g.*, a detectably-labelled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention; or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention.

[115] The kit can also comprise, *e.g.*, a buffering agent, a preservative or a protein-stabilizing agent. The kit can further comprise components necessary for detecting the detectable-label, *e.g.*, an enzyme or a substrate. The kit can also contain a control sample or a series of control samples, which can be assayed and compared to the test sample. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[116] *Making Polymorphisms and Mutations of the Invention.* Effects of the polymorphisms and mutations identified herein on gene expression may be investigated by preparing recombinant cells and/or organisms, preferably recombinant animals, containing a polymorphic variant and/or mutation of the gene.

[117] In one aspect, the present invention includes one or more polynucleotides encoding mutant or polymorphic polypeptides, including degenerate variants thereof. The invention also encompasses allelic variants of the same, that is, naturally occurring alternative forms of the isolated polynucleotides that encode mutant polypeptides that are identical, homologous or related to those encoded by the polynucleotides. Alternatively, non-naturally occurring

variants may be produced by mutagenesis techniques or by direct synthesis techniques well known in the art. Accordingly, nucleic acid sequences capable of hybridizing at low stringency with any nucleic acid sequences encoding mutant polypeptide of the present invention are considered to be within the scope of the invention. For example, for a nucleic acid sequence of about 20-40 bases, a typical prehybridization, hybridization, and wash protocol is as follows: (1) prehybridization: incubate nitrocellulose filters containing the denatured target DNA for 3-4 hours at 55°C in 5xDenhardt's solution, 6xSSC (20xSSC consists of 175 g NaCl, 88.2 g sodium citrate in 800 ml H_2O adjusted to pH. 7.0 with 10 N NaOH), 0.1% SDS, and 100 mg/ml denatured salmon sperm DNA, (2) hybridization: incubate filters in prehybridization solution plus probe at 42°C for 14-48 hours, (3) wash; three 15 minutes washes in 6xSSC and 0.1% SDS at room temperature, followed by a final 1-1.5 minutes wash in 6xSSC and 0.1% SDS at 55°C. Other equivalent procedures, *e.g.*, employing organic solvents such as formamide, are well known in the art. Standard stringency conditions are well characterized in standard molecular biology cloning texts. See, for example, Sambrook, Fritsch, & Maniatis, *Molecular Cloning A Laboratory Manual, 2nd Ed.*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbour, New York, 1989); Glover DN, *DNA Cloning, Volumes I and II* , (1985); *Oligonucleotide Synthesis,* Gait MJ, ed. (1984); *Nucleic Acid Hybridization,* Hames BD & Higgins SJ, eds. (1984).

[118] *Recombinant Expression Vectors.* Another aspect of the invention includes vectors containing one or more nucleic acid sequences encoding a mutant or polymorphic polypeptide. In practicing the present invention, many conventional techniques in molecular biology, microbiology and recombinant DNA are used. These techniques are well known and are explained in, *e.g.*, *Current Protocols in Molecular Biology, Vols. I-III,* Ausubel, ed. (1997); Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Edition.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989); Glover DN, *DNA Cloning: A Practical Approach, Vols. I and II* (1985); *Oligonucleotide Synthesis*, Gait, Ed. (1984); *Nucleic Acid Hybridization*, Hames & Higgins, eds. (1985); *Transcription and Translation*, Hames & Higgins, Eds. (1984); *Animal Cell* Culture, Freshney, ed. (1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); Perbal, *A Practical Guide to Molecular Cloning*; the series *Methods in Enzymol.*, (Academic Press, Inc., 1984); *Gene Transfer Vectors for Mammalian Cells*, Miller & Calos, eds. (Cold Spring Harbor Press, Cold Spring Harbor Laboratory, New

York, 1987); and *Methods in Enzymology, Vols. 154 and 155*, Wu & Grossman, and Wu, Eds., respectively.

[119] For recombinant expression of one or more the polypeptides of the invention, the nucleic acid containing all or a portion of the nucleotide sequence encoding the polypeptide is inserted into an appropriate cloning vector, or an expression vector (*i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted polypeptide coding sequence) by recombinant DNA techniques well known in the art and as detailed below.

[120] In general, expression vectors useful in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors that are not technically plasmids, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions. Such viral vectors permit infection of a subject and expression in that subject of a compound. Becker *et al.*, *Meth. Cell Biol.* 43: 161 89 (1994).

[121] The recombinant expression vectors of the invention comprise a nucleic acid encoding a mutant or polymorphic polypeptide in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequences in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

[122] The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed,

the level of expression of polypeptide desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce polypeptides or peptides, including fusion polypeptides, encoded by nucleic acids as described herein (*e.g.*, mutant polypeptides and mutant-derived fusion polypeptides, etc.).

[123] *Mutant and Polymorphic Polypeptide-Expressing Host Cells.* Another aspect of the invention pertains to mutant and polymorphic polypeptide-expressing host cells, which contain a nucleic acid encoding one or more mutant/polymorphic polypeptides of the invention. To prepare a recombinant cell of the invention, the desired isogene may be introduced into a host cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the isogene is introduced into a cell in such a way that it recombines with the endogenous gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired gene polymorphism or mutation. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner.

[124] The recombinant expression vectors of the invention can be designed for expression of mutant polypeptides in prokaryotic or eukaryotic cells. For example, mutant/polymorphic polypeptides can be expressed in bacterial cells such as *Escherichia coli* (*E. coli*), insect cells (using baculovirus expression vectors), fungal cells, *e.g.*, yeast, yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase. The SMP2 promoter is useful in the expression of polypeptides in smooth muscle cells (Qian *et al.*, Endocrinology 140(4): 1826 (1999)).

[125] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide

encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant polypeptide; (ii) to increase the solubility of the recombinant polypeptide; and (iii) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, *Gene* 67: 31 40 (1988)), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[126] Examples of suitable inducible non fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, Gene 69:301 315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990) pp. 60-89).

[127] One strategy to maximize recombinant polypeptide expression in *E. coli* is to express the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide. See, *e.g.*, Gottesman, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990) 119 128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the expression host, *e.g.*, *E. coli* (see, *e.g.*, Wada *et al.*, *Nucl. Acids Res.* 20: 2111-2118 (1992)). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques. In another embodiment, the mutant/polymorphic polypeptide expression vector is a yeast expression vector.

[128] Examples of vectors for expression in yeast *Saccharomyces cerivisiae* include pYepSec1 (Baldari *et al.*, *EMBO J.* 6: 229 234 (1987)), pMFa (Kurjan & Herskowitz, Cell 30: 933 943 (1982)), pJRY88 (Schultz *et al.*, *Gene* 54: 113 123 (1987)), pYES2 (InVitrogen Corporation, San Diego, Calif., USA), and picZ (InVitrogen Corp, San Diego, Calif., USA). Alternatively, mutant polypeptide can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of polypeptides in cultured

insect cells (*e.g.*, SF9 cells) include the pAc series (Smith *et al.*, *Mol. Cell. Biol.* 3: 2156 2165 (1983)) and the pVL series (Lucklow & Summers, *Virology* 170: 31 39 (1989)).

[129] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, *Nature* 329: 842 846 (1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6: 187 195 (1987)). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed.*(Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).

[130] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue specific regulatory elements are used to express the nucleic acid). Tissue specific regulatory elements are known in the art. Nonlimiting examples of suitable tissue specific promoters include the albumin promoter (liver specific; Pinkert, *et al.*, *Genes Dev.* 1: 268 277 (1987)), lymphoid specific promoters (Calame & Eaton, *Adv. Immunol.* 43: 235 275 (1988)), in particular promoters of T cell receptors (Winoto & Baltimore, *EMBO J.* 8: 729 733 (1989)) and immunoglobulins (Banerji *et al.*, *Cell* 33: 729 740 (1983); Queen & Baltimore, *Cell* 33: 741 748 (1983)), neuron specific promoters (*e.g.*, the neurofilament promoter; Byrne & Ruddle, *Proc. Natl. Acad. Sci. USA* 86: 5473 5477 (1989)), pancreas specific promoters (Edlund *et al.*, *Science* 230: 912 916 (1985)), and mammary gland specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally regulated promoters are also encompassed, *e.g.*, the murine *hox* promoters (Kessel & Gruss, *Science* 249: 374 379 (1990)) and the α-fetoprotein promoter (Campes & Tilghman, *Genes Dev.* 3: 537 546 (1989)).

[131] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to a mutant polypeptide mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the

antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see, *e.g.*, Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews Trends in Genetics, Vol. 1(1)* (1986).

[132] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[133] A host cell can be any prokaryotic or eukaryotic cell. For example, mutant polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[134] Vector DNA can be introduced into prokaryotic or eukaryotic cells *via* conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co precipitation, DEAE dextran mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd ed.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989), and other laboratory manuals.

[135] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that

confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding mutant polypeptide or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

[136] A host cell that includes a compound of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) recombinant mutant/polymorphic polypeptide. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding mutant/polymorphic polypeptide has been introduced) in a suitable medium such that mutant polypeptide is produced. In another embodiment, the method further comprises the step of isolating mutant/polymorphic polypeptide from the medium or the host cell. Purification of recombinant polypeptides is well known in the art and includes ion exchange purification techniques, or affinity purification techniques, for example with an antibody to the compound. Methods of creating antibodies to the compounds of the present invention are discussed below.

[137] *Transgenic Animals.* Recombinant organisms, *i.e.*, transgenic animals, expressing a variant gene of the invention are prepared using standard procedures known in the art. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. S*ee, e.g.*, U.S. Pat. No. 5,610,053 and "The Introduction of Foreign Genes into Mice" and the cited references therein, in: *Recombinant DNA*, Eds. Watson JD, Gilman M, Witkowski J & Zoller M (W.H. Freeman and Company, New York) pp. 254-272. Transgenic animals stably expressing a human isogene and producing human protein can be used as biological models for studying diseases related to abnormal expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

[138] *Characterizing Gene Expression Level.* Methods to detect and measure mRNA levels (*i.e.*, gene transcription level) and levels of polypeptide gene expression products (*i.e.*, gene translation level) are well-known in the art and include the use of nucleotide microarrays and polypeptide detection methods involving mass spectrometers, reverse-transcription and amplification and/or antibody detection and quantification techniques. See also, Tom

Strachan & Andrew Read, *Human Molecular Genetics, 2nd Edition.* (John Wiley and Sons, Inc. Publication, New York, 1999)).

[139] *Determination of Target Gene Transcription.* The determination of the level of the expression product of the gene in a biological sample, *e.g.*, the tissue or body fluids of an individual, may be performed in a variety of ways. The term "biological sample" is intended to include tissues, cells, biological fluids and isolates thereof, isolated from a subject, as well as tissues, cells and fluids present within a subject. Many expression detection methods use isolated RNA. For *in vitro* methods, any RNA isolation technique that does not select against the isolation of mRNA can be utilized for the purification of RNA from cells. See, *e.g.*, Ausubel *et al.*, Ed., *Curr. Prot. Mol. Biol.* (John Wiley & Sons, New York, 1987-1999).

[140] In one embodiment, the level of the mRNA expression product of the target gene is determined. Methods to measure the level of a specific mRNA are well-known in the art and include Northern blot analysis, reverse transcription PCR and real time quantitative PCR or by hybridization to a oligonucleotide array or microarray. In other more preferred embodiments, the determination of the level of expression may be performed by determination of the level of the protein or polypeptide expression product of the gene in body fluids or tissue samples including but not limited to blood or serum. Large numbers of tissue samples can readily be processed using techniques well-known to those of skill in the art, such as, *e.g.*, the single-step RNA isolation process of U.S. Pat. No. 4,843,155.

[141] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, PCR analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, *e.g.*, a full-length cDNA, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to an mRNA or genomic DNA encoding a marker of the present invention. Other suitable probes for use in the diagnostic assays of the invention are described herein. Hybridization of an mRNA with the probe indicates that the marker in question is being expressed.

[142] In one format, the probes are immobilized on a solid surface and the mRNA is contacted with the probes, for example, in an Affymetrix gene chip array (Affymetrix, Calif.

USA). A skilled artisan can readily adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the markers of the present invention.

[143] An alternative method for determining the level of mRNA corresponding to a marker of the present invention in a sample involves the process of nucleic acid amplification, *e.g.*, by RT-PCR (the experimental embodiment set forth in U.S. Pat. No. 4,683,202); ligase chain reaction (Barany *et al., Proc. Natl. Acad. Sci. USA* 88:189-193 (1991)) self-sustained sequence replication (Guatelli *et al., Proc. Natl. Acad. Sci. USA* 87: 1874-1878 (1990)); transcriptional amplification system (Kwoh *et al., Proc. Natl. Acad. Sci. USA* 86: 1173-1177 (1989)); Q-Beta Replicase (Lizardi *et al., Biol. Technology* 6: 1197 (1988)); rolling circle replication (U.S. Pat. No. 5,854,033); or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially useful for the detection of the nucleic acid molecules if such molecules are present in very low numbers. As used herein, "amplification primers" are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10-30 nucleotides in length and flank a region from about 50-200 nucleotides in length.

[144] Real-time quantitative PCR (RT-PCR) is one way to assess gene expression levels, *e.g.*, of genes of the invention, *e.g.*, those containing SNPs and mutations of interest. The RT-PCR assay utilizes an RNA reverse transcriptase to catalyze the synthesis of a DNA strand from an RNA strand, including an mRNA strand. The resultant DNA may be specifically detected and quantified and this process may be used to determine the levels of specific species of mRNA. One method for doing this is TAQMAN® (PE Applied Biosystems, Foster City, Calif., USA) and exploits the 5' nuclease activity of AMPLITAQ GOLD™ DNA polymerase to cleave a specific form of probe during a PCR reaction. This is referred to as a TAQMAN™ probe. See Luthra *et al., Am. J. Pathol.* 153: 63-68 (1998); Kuimelis *et al., Nucl. Acids Symp. Ser.* 37: 255-256 (1997); and Mullah *et al., Nucl. Acids Res.* 26(4): 1026-1031 (1998)). During the reaction, cleavage of the probe separates a reporter dye and a quencher dye, resulting in increased fluorescence of the reporter. The accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the reporter dye. Heid *et al., Genome Res.* 6(6): 986-994 (1996)). The higher the starting copy number of

nucleic acid target, the sooner a significant increase in fluorescence is observed. See Gibson, Heid & Williams *et al.*, *Genome Res.* 6: 995-1001 (1996).

[145] Other technologies for measuring the transcriptional state of a cell produce pools of restriction fragments of limited complexity for electrophoretic analysis, such as methods combining double restriction enzyme digestion with phasing primers (see, *e.g.*, EP 0 534858 A1), or methods selecting restriction fragments with sites closest to a defined mRNA end. (See, *e.g.*, Prashar & Weissman, *Proc. Natl. Acad. Sci. USA* 93(2) 659-663 (1996)).

[146] Other methods statistically sample cDNA pools, such as by sequencing sufficient bases, *e.g.*, 20-50 bases, in each of multiple cDNAs to identify each cDNA, or by sequencing short tags, *e.g.*, 9-10 bases, which are generated at known positions relative to a defined mRNA end pathway pattern. See, *e.g.*, Velculescu, *Science* 270: 484-487 (1995). The cDNA levels in the samples are quantified and the mean, average and standard deviation of each cDNA is determined using by standard statistical means well-known to those of skill in the art. Norman T.J. Bailey, *Statistical Methods In Biology, 3rd Edition* (Cambridge University Press, 1995).

[147] *Detection of Polypeptides. Immunological Detection Methods.* Expression of the protein encoded by the genes of the invention can be detected by a probe which is detectably labelled, or which can be subsequently labelled. The term "labelled", with regard to the probe or antibody, is intended to encompass direct-labelling of the probe or antibody by coupling, *i.e.*, physically linking, a detectable substance to the probe or antibody, as well as indirect-labelling of the probe or antibody by reactivity with another reagent that is directly-labelled. Examples of indirect labelling include detection of a primary antibody using a fluorescently-labelled secondary antibody and end-labelling of a DNA probe with biotin such that it can be detected with fluorescently-labelled streptavidin. Generally, the probe is an antibody that recognizes the expressed protein. A variety of formats can be employed to determine whether a sample contains a target protein that binds to a given antibody. Immunoassay methods useful in the detection of target polypeptides of the present invention include, but are not limited to, *e.g.*, dot blotting, western blotting, protein chips, competitive and non-competitive protein binding assays, enzyme-linked immunosorbant assays (ELISA), immunohistochemistry, fluorescence activated cell sorting (FACS), and others commonly used and widely-described in scientific and patent literature, and many employed commercially. A skilled artisan can readily adapt known protein/antibody detection methods

for use in determining whether cells express a marker of the present invention and the relative concentration of that specific polypeptide expression product in blood or other body tissues. Proteins from individuals can be isolated using techniques that are well-known to those of skill in the art. The protein isolation methods employed can, *e.g.*, be such as those described in Harlow & Lane, *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988)).

[148] For the production of antibodies to a protein encoded by one of the disclosed genes, various host animals may be immunized by injection with the polypeptide, or a portion thereof. Such host animals may include, but are not limited to, rabbits, mice and rats. Various adjuvants may be used to increase the immunological response, depending on the host species including, but not limited to, Freund's (complete and incomplete), mineral gels, such as aluminium hydroxide; surface active substances, such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin and dinitrophenol; and potentially useful human adjuvants, such as bacille Camette-Guerin (BCG) and *Corynebacterium parvum.*

[149] Monoclonal antibodies (mAbs), which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler & Milstein, *Nature* 256: 495-497 (1975); and U.S. Pat. No. 4,376,110; the human B-cell hybridoma technique of Kosbor *et al.*, *Immunol. Today* 4: 72 (1983); Cole *et al., Proc. Natl. Acad. Sci. USA* 80: 2026-2030 (1983); and the EBV-hybridoma technique of Cole *et al., Monoclonal Antibodies and Cancer Therapy* (Alan R. Liss, Inc., 1985) pp. 77-96.

[150] In addition, techniques developed for the production of "chimeric antibodies" (see Morrison *et al., Proc. Natl. Acad. Sci. USA* 81: 6851-6855 (1984); Neuberger *et al., Nature* 312: 604-608 (1984); and Takeda *et al., Nature* 314: 452-454 (1985)), by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable or hypervariable region derived form a murine mAb and a human immunoglobulin constant region.

[151] Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988); and Ward *et al.*, *Nature* 334: 544-546 (1989)) can be adapted to produce differentially expressed gene single-chain antibodies.

[152] Techniques useful for the production of "humanized antibodies" can be adapted to produce antibodies to the proteins, fragments or derivatives thereof. Such techniques are disclosed in U.S. Pat. Nos. 5,932,448; 5,693,762; 5,693,761; 5,585,089; 5,530,101; 5,569,825; 5,625,126; 5,633,425; 5,789,650; 5,661,016; and 5,770,429.

[153] Antibodies or antibody fragments can be used in methods, such as Western blots or immunofluorescence techniques, to detect the expressed proteins. In such uses, it is generally preferable to immobilize either the antibody or proteins on a solid support. Suitable solid phase supports or carriers include any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros and magnetite.

[154] A useful method, for ease of detection, is the sandwich ELISA, of which a number of variations exist, all of which are intended to be used in the methods and assays of the present invention. As used herein, "sandwich assay" is intended to encompass all variations on the basic two-site technique. Immunofluorescence and EIA techniques are both very well-established in the art. However, other reporter molecules, such as radioisotopes, chemiluminescent or bioluminescent molecules may also be employed. It will be readily apparent to the skilled artisan how to vary the procedure to suit the required use.

[155] Whole genome monitoring of protein, *i.e.*, the "proteome," can be carried out by constructing a microarray in which binding sites comprise immobilized, preferably monoclonal, antibodies specific to a plurality of protein species encoded by the cell genome. Preferably, antibodies are present for a substantial fraction of the encoded proteins, or at least for those proteins relevant to testing or confirming a biological network model of interest. As noted above, methods for making monoclonal antibodies are well-known. See, *e.g.*, Harlow & Lane, *Antibodies: A Laboratory Manual*" (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988)). In a preferred embodiment, monoclonal antibodies are raised against synthetic peptide fragments designed based on genomic sequence of the cell.

With such an antibody array, proteins from the cell are contacted to the array and their binding is measured with assays known in the art.

[156] *Detection of Polypeptides. Two-Dimensional Gel Electrophoresis.* Two-dimensional gel electrophoresis is well-known in the art and typically involves isoelectric focusing along a first dimension followed by SDS-PAGE electrophoresis along a second dimension. See, *e.g.*, Hames *et al.*, *Gel Electrophoresis of Proteins: A Practical Approach* (IRL Press, New York, 1990); Shevchenko *et al.*, *Proc. Natl. Acad. Sci. USA* 93: 14440-14445 (1996); Sagliocco *et al.*, *Yeast* 12: 1519-1533 (1996); and Lander, *Science* 274: 536-539 (1996)).

[157] *Detection of Polypeptides. Mass Spectroscopy.* The identity as well as expression level of target polypeptide can be determined using mass spectrocopy technique (MS). MS-based analysis methodology is useful for analysis of isolated target polypeptide as well as analysis of target polypeptide in a biological sample. MS formats for use in analyzing a target polypeptide include ionization (I) techniques, such as, but not limited to, matrix assisted laser desorption (MALDI), continuous or pulsed electrospray ionization (ESI) and related methods, such as ionspray or thermospray, and massive cluster impact (MCI). Such ion sources can be matched with detection formats, including linear or non-linear reflectron time of flight (TOF), single or multiple quadrupole, single or multiple magnetic sector Fourier transform ion cyclotron resonance (FTICR), ion trap and combinations thereof such as ion-trap/TOF. For ionization, numerous matrix/wavelength combinations (*e.g.*, matrix assisted laser desorption (MALDI)) or solvent combinations (*e.g.*, ESI) can be employed.

[158] For mass spectroscopy (MS) analysis, the target polypeptide can be solubilised in an appropriate solution or reagent system. The selection of a solution or reagent system, *e.g.*, an organic or inorganic solvent, will depend on the properties of the target polypeptide and the type of MS performed, and is based on methods well-known in the art. See, *e.g.*, Vorm *et al.*, *Anal. Chem.* 61: 3281 (1994) for MALDI; and Valaskovic *et al.*, *Anal. Chem.* 67: 3802 (1995), for ESI. MS of peptides also is described, *e.g.*, in International PCT Application No. WO 93/24834 and U.S. Pat. No. 5,792,664. A solvent is selected that minimizes the risk that the target polypeptide will be decomposed by the energy introduced for the vaporization process. A reduced risk of target polypeptide decomposition can be achieved, *e.g.*, by embedding the sample in a matrix. A suitable matrix can be an organic compound such as a sugar, *e.g.*, a pentose or hexose, or a polysaccharide such as cellulose. Such compounds are decomposed thermolytically into CO_2 and H_2O such that no residues are formed that can lead

to chemical reactions. The matrix also can be an inorganic compound, such as nitrate of ammonium, which is decomposed essentially without leaving any residue. Use of these and other solvents is known to those of skill in the art. See, *e.g.*, U.S. Pat. No. 5,062,935. Electrospray MS has been described by Fenn *et al.*, *J. Phys. Chem.* 88: 4451-4459 (1984); and PCT Application No. WO 90/14148; and current applications are summarized in review articles. See Smith *et al.*, *Anal. Chem.* 62: 882-89 (1990); and Ardrey, *Spectroscopy* 4: 10-18 (1992).

[159] The mass of a target polypeptide determined by MS can be compared to the mass of a corresponding known polypeptide. For example, where the target polypeptide is a mutant protein, the corresponding known polypeptide can be the corresponding non-mutant protein, *e.g.*, wild-type protein. With ESI, the determination of molecular weights in femtomole amounts of sample is very accurate due to the presence of multiple ion peaks, all of which can be used for mass calculation. Sub-attomole levels of protein have been detected, *e.g.*, using ESI MS (Valaskovic *et al.*, *Science* 273: 1199-1202 (1996)) and MALDI MS (Li *et al.*, *J. Am. Chem. Soc.* 118: 1662-1663 (1996)).

[160] *Matrix Assisted Laser Desorption (MALDI).* The level of the target protein in a biological sample, *e.g.*, body fluid or tissue sample, may be measured by means of mass spectrometric (MS) methods including, but not limited to, those techniques known in the art as matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry (MALDI-TOF-MS) and surfaces enhanced for laser desorption/ionization, time-of-flight mass spectrometry (SELDI-TOF-MS) as further detailed below. Methods for performing MALDI are well-known to those of skill in the art. See, *e.g.*, Juhasz *et al.*, *Analysis, Anal. Chem.* 68: 941-946 (1996), and see also, *e.g.*, U.S. Pat. Nos. 5,777,325; 5,742,049; 5,654,545; 5,641,959; 5,654,545 and 5,760,393 for descriptions of MALDI and delayed extraction protocols. Numerous methods for improving resolution are also known. MALDI-TOF-MS has been described by Hillenkamp *et al.*, *Biological Mass Spectrometry*, Burlingame & McCloskey, eds. (Elsevier Science Publ., Amsterdam, 1990) pp. 49-60.

[161] A variety of techniques for marker detection using mass spectroscopy can be used. See *Bordeaux Mass Spectrometry Conference Report*, Hillenkamp, Ed., pp. 354-362 (1988); *Bordeaux Mass Spectrometry Conference Report*, Karas & Hillenkamp, Eds., pp. 416-417 (1988); Karas & Hillenkamp, *Anal. Chem.* 60: 2299-2301 (1988); and Karas *et al.*, *Biomed. Environ. Mass Spectrum* 18: 841-843 (1989). The use of laser beams in TOF-MS is shown,

e.g., in U.S. Patent Nos. 4,694,167; 4,686,366, 4,295,046 and 5,045,694, which are incorporated herein by reference in their entireties. Other MS techniques allow the successful volatilization of high molecular weight biopolymers, without fragmentation, and have enabled a wide variety of biological macromolecules to be analyzed by mass spectrometry.

[162] *Surfaces Enhanced for Laser Desorption/Ionization (SELDI).* Other techniques are used which employ new MS probe element compositions with surfaces that allow the probe element to actively participate in the capture and docking of specific analytes, described as Affinity Mass Spectrometry (AMS). See SELDI patents U.S. Pat. Nos. 5,719,060; 5,894,063; 6,020,208; 6,027,942; 6,124,137; and U.S. Patent application No. U.S. 2003/0003465. Several types of new MS probe elements have been designed with Surfaces Enhanced for Affinity Capture (SEAC). See Hutchens & Yip, *Rapid Commun. Mass Spectrom.* 7: 576-580 (1993). SEAC probe elements have been used successfully to retrieve and tether different classes of biopolymers, particularly proteins, by exploiting what is known about protein surface structures and biospecific molecular recognition. The immobilized affinity capture devices on the MS probe element surface, *i.e.*, SEAC, determines the location and affinity (specificity) of the analyte for the probe surface, therefore the subsequent analytical MS process is efficient.

[163] Within the general category of SELDI are three separate subcategories: (1) Surfaces Enhanced for Neat Desorption (SEND), where the probe element surfaces, *i.e.*, sample presenting means, are designed to contain Energy Absorbing Molecules (EAM) instead of "matrix" to facilitate desorption/ionizations of analytes added directly (neat) to the surface. (2) SEAC, where the probe element surfaces, *i.e.*, sample presenting means, are designed to contain chemically defined and/or biologically defined affinity capture devices to facilitate either the specific or non-specific attachment or adsorption (so-called docking or tethering) of analytes to the probe surface, by a variety of mechanisms (mostly non-covalent). (3) Surfaces Enhanced for Photolabile Attachment and Release (SEPAR), where the probe element surfaces, *i.e.*, sample presenting means, are designed or modified to contain one or more types of chemically defined cross-linking molecules to serve as covalent docking devices. The chemical specificities determining the type and number of the photolabile molecule attachment points between the SEPAR sample presenting means (*i.e.*, probe element surface) and the analyte (*e.g.*, protein) may involve any one or more of a number of different residues

or chemical structures in the analyte (*e.g.*, His, Lys, Arg, Tyr, Phe and Cys residues in the case of proteins and peptides).

[164] *Functionalizing Polypeptides.* A polypeptide of interest also can be modified to facilitate conjugation to a solid support. A chemical or physical moiety can be incorporate into the polypeptide at an appropriate position. For example, a polypeptide of interest can be modified by adding an appropriate functional group to the carboxyl terminus or amino terminus of the polypeptide, or to an amino acid in the peptide, (*e.g.*, to a reactive side chain, or to the peptide backbone. The artisan will recognize, however, that such a modification, *e.g.*, the incorporation of a biotin moiety, can affect the ability of a particular reagent to interact specifically with the polypeptide and, accordingly, will consider this factor, if relevant, in selecting how best to modify a polypeptide of interest. A naturally-occurring amino acid normally present in the polypeptide also can contain a functional group suitable for conjugating the polypeptide to the solid support. For example, a cysteine residue present in the polypeptide can be used to conjugate the polypeptide to a support containing a sulfhydryl group through a disulfide linkage, *e.g.*, a support having cysteine residues attached thereto. Other bonds that can be formed between two amino acids, include, but are not limited to, *e.g.*, monosulfide bonds between two lanthionine residues, which are non-naturally-occurring amino acids that can be incorporated into a polypeptide; a lactam bond formed by a transamidation reaction between the side chains of an acidic amino acid and a basic amino acid, such as between the y-carboxyl group of Glu (or alpha carboxyl group of Asp) and the amino group of Lys; or a lactone bond produced, *e.g.*, by a crosslink between the hydroxy group of Ser and the carboxyl group of Glu (or alpha carboxyl group of Asp). Thus, a solid support can be modified to contain a desired amino acid residue, *e.g.*, a Glu residue, and a polypeptide having a Ser residue, particularly a Ser residue at the N-terminus or C-terminus, can be conjugated to the solid support through the formation of a lactone bond. The support need not be modified to contain the particular amino acid, *e.g.*, Glu, where it is desired to form a lactone-like bond with a Ser in the polypeptide, but can be modified, instead, to contain an accessible carboxyl group, thus providing a function corresponding to the alpha carboxyl group of Glu.

[165] *Thiol-Reactive Functionalities.* A thiol-reactive functionality is particularly useful for conjugating a polypeptide to a solid support. A thiol-reactive functionality is a chemical group that can rapidly react with a nucleophilic thiol moiety to produce a covalent bond, *e.g.*,

a disulfide bond or a thioether bond. A variety of thiol-reactive functionalities are known in the art, including, *e.g.*, haloacetyls, such as iodoacetyl; diazoketones; epoxy ketones, alpha- and beta-unsaturated carbonyls, such as alpha-enones and beta-enones; and other reactive Michael acceptors, such as maleimide; acid halides; benzyl halides; and the like. See Greene & Wuts, *Protective Groups in Organic Synthesis*, 2nd Edition (John Wiley & Sons, 1991).

[166] If desired, the thiol groups can be blocked with a photocleavable protecting group, which then can be selectively cleaved, *e.g.*, by photolithography, to provide portions of a surface activated for immobilization of a polypeptide of interest. Photocleavable protecting groups are known in the art (see, *e.g.*, published International PCT Application No. WO 92/10092; and McCray *et al.*, *Ann. Rev. Biophys. Biophys. Chem.* 18: 239-270 (1989)) and can be selectively de-blocked by irradiation of selected areas of the surface using, *e.g.*, a photolithography mask.

[167] *Linkers.* A polypeptide of interest can be attached directly to a support *via* a linker. Any linkers known to those of skill in the art to be suitable for linking peptides or amino acids to supports, either directly or *via* a spacer, may be used. For example, the polypeptide can be conjugated to a support, such as a bead, through means of a variable spacer. Linkers, include, Rink amide linkers (see, *e.g.*, Rink, *Tetrahedron Lett.* 28: 3787 (1976)); trityl chloride linkers (see, *e.g.*, Leznoff, *Ace Chem. Res.* 11: 327 (1978)); and Merrifield linkers (see, *e.g.*, Bodansky *et al.*, *Peptide Synthesis,* 2nd Edition (Academic Press, New York, 1976)). For example, trityl linkers are known. See, *e.g.*, U.S. Pat. Nos. 5,410,068 and 5,612,474. Amino trityl linkers are also known. See, *e.g.*, U.S. Pat. No. 5,198,531. Other linkers include those that can be incorporated into fusion proteins and expressed in a host cell. Such linkers may be selected amino acids, enzyme substrates or any suitable peptide. The linker may be made, *e.g.*, by appropriate selection of primers when isolating the nucleic acid. Alternatively, they may be added by post-translational modification of the protein of interest. Linkers that are suitable for chemically linking peptides to supports, include disulfide bonds, thioether bonds, hindered disulfide bonds and covalent bonds between free reactive groups, such as amine and thiol groups.

[168] *Cleavable Linkers.* A linker can provide a reversible linkage such that it is cleaved under the select conditions. In particular, selectively cleavable linkers, including photocleavable linkers (see U.S. Pat. No. 5,643,722), acid cleavable linkers (see Fattom *et al.*, *Infect. Immun.* 60: 584-589 (1992)), acid-labile linkers (see Welhöner *et al.*, *J. Biol. Chem.*

266: 4309-4314 (1991)) and heat sensitive linkers are useful. A linkage can be, *e.g.*, a disulfide bond, which is chemically cleavable by mercaptoethanol or dithioerythrol; a biotin/streptavidin linkage, which can be photocleavable; a heterobifunctional derivative of a trityl ether group, which can be cleaved by exposure to acidic conditions or under conditions of MS (see Köster *et al*, *Tetrahedron Lett.* 31: 7095 (1990)); a levulinyl-mediated linkage, which can be cleaved under almost neutral conditions with a hydrazinium/acetate buffer; an arginine-arginine or a lysine-lysine bond, either of which can be cleaved by an endopeptidase, such as trypsin; a pyrophosphate bond, which can be cleaved by a pyrophosphatase; or a ribonucleotide bond, which can be cleaved using a ribonuclease or by exposure to alkali condition. A photolabile cross-linker, such as 3-amino-(2-nitrophenyl)propionic acid can be employed as a means for cleaving a polypeptide from a solid support. Brown *et al.*, *Mol Divers*, pp. 4-12 (1995); Rothschild *et al.*, *Nucl. Acids. Res.* 24: 351-66 (1996); and U.S. Pat. No. 5,643,722. Other linkers include RNA linkers that are cleavable by ribozymes and other RNA enzymes and linkers, such as the various domains, such as CH_1, CH_2 and CH_3, from the constant region of human IgG1. See, Batra *et al.*, *Mol Immunol* 30: 379-396 (1993).

[169] Combinations of any linkers are also contemplated herein. For example, a linker that is cleavable under MS conditions, such as a silyl linkage or photocleavable linkage, can be combined with a linker, such as an avidin biotin linkage, that is not cleaved under these conditions, but may be cleaved under other conditions. Acid-labile linkers are particularly useful chemically cleavable linkers for mass spectrometry, especially for MALDI-TOF, because the acid labile bond is cleaved during conditioning of the target polypeptide upon addition of a 3-HPA matrix solution. The acid labile bond can be introduced as a separate linker group, *e.g.*, an acid labile trityl group, or can be incorporated in a synthetic linker by introducing one or more silyl bridges using diisopropylysilyl, thereby forming a diisopropylysilyl linkage between the polypeptide and the solid support. The diisopropylysilyl linkage can be cleaved using mildly acidic conditions, such as 1.5% trifluoroacetic acid (TFA) or 3-HPA/1 % TFA MALDI-TOF matrix solution. Methods for the preparation of diisopropylysilyl linkages and analogues thereof are well-known in the art. See, *e.g.*, Saha *et al.*, *J. Org. Chem.* 58: 7827-7831 (1993).

[170] *Use of a Pin Tool to Immobilize a Polypeptide.* The immobilization of a polypeptide of interest to a solid support using a pin tool can be particularly advantageous. Pin tools include

those disclosed herein or otherwise known in the art. See, *e.g.*, U.S. Application Serial Nos. 08/786,988 and 08/787,639; and International PCT Application No. WO 98/20166.

[171] A pin tool in an array, *e.g.*, a 4 × 4 array, can be applied to wells containing polypeptides of interest. Where the pin tool has a functional group attached to each pin tip, or a solid support, *e.g.*, functionalized beads or paramagnetic beads are attached to each pin, the polypeptides in a well can be captured (1 pmol capacity). During the capture step, the pins can be kept in motion (vertical, 1-2 mm travel) to increase the efficiency of the capture. Where a reaction, such as an *in vitro* transcription is being performed in the wells, movement of the pins can increase efficiency of the reaction. Further immobilization can result by applying an electrical field to the pin tool. When a voltage is applied to the pin tool, the polypeptides are attracted to the anode or the cathode, depending on their net charge.

[172] For more specificity, the pin tool (with or without voltage) can be modified to have conjugated thereto a reagent specific for the polypeptide of interest, such that only the polypeptides of interest are bound by the pins. For example, the pins can have nickel ions attached, such that only polypeptides containing a polyhistidine sequence are bound. Similarly, the pins can have antibodies specific for a target polypeptide attached thereto, or to beads that, in turn, are attached to the pins, such that only the target polypeptides, which contain the epitope recognized by the antibody, are bound by the pins.

[173] Captured polypeptides can be analyzed by a variety of means including, *e.g.*, spectrometric techniques, such as UV/VIS, IR, fluorescence, chemiluminescence, NMR spectroscopy, MS or other methods known in the art, or combinations thereof. If conditions preclude direct analysis of captured polypeptides, the polypeptides can be released or transferred from the pins, under conditions such that the advantages of sample concentration are not lost. Accordingly, the polypeptides can be removed from the pins using a minimal volume of eluent, and without any loss of sample. Where the polypeptides are bound to the beads attached to the pins, the beads containing the polypeptides can be removed from the pins and measurements made directly from the beads.

[174] Pin tools can be useful for immobilizing polypeptides of interest in spatially addressable manner on an array. Such spatially addressable or pre-addressable arrays are useful in a variety of processes, including, for example, quality control and amino acid sequencing diagnostics. The pin tools described in the U.S. Application Nos. 08/786,988 and 08/787,639 and International PCT Application No. WO 98/20166 are serial and parallel

dispensing tools that can be employed to generate multi-element arrays of polypeptides on a surface of the solid support. The array surface can be flat, with beads or geometrically altered to include wells, which can contain beads. In addition, MS geometries can be adapted for accommodating a pin tool apparatus.

[175] *Other Aspects of the Biological State.* In various embodiments of the invention, aspects of the biological activity state, or mixed aspects can be measured in order to obtain drug and pathway responses. The activities of proteins relevant to the characterization of cell function can be measured, and embodiments of this invention can be based on such measurements. Activity measurements can be performed by any functional, biochemical or physical means appropriate to the particular activity being characterized. Where the activity involves a chemical transformation, the cellular protein can be contacted with natural substrates, and the rate of transformation measured. Where the activity involves association in multimeric units, *e.g.*, association of an activated DNA binding complex with DNA, the amount of associated protein or secondary consequences of the association, such as amounts of mRNA transcribed, can be measured. Also, where only a functional activity is known, *e.g.*, as in cell cycle control, performance of the function can be observed. However known and measured, the changes in protein activities form the response data analyzed by the methods of this invention. In alternative and non-limiting embodiments, response data may be formed of mixed aspects of the biological state of a cell. Response data can be constructed from, *e.g.*, changes in certain mRNA abundances, changes in certain protein abundances and changes in certain protein activities.

[176] The following EXAMPLES are presented in order to more fully illustrate the preferred embodiments of the invention. These EXAMPLES should in no way be construed as limiting the scope of the invention, as defined by the appended claims.

EXAMPLE I

IDENTIFICATION OF EGFR MUTATIONS IN HUMAN COLORECTAL CANCER AND HUMAN LUNG CANCER

[177] DHPLC analysis (Lilleberg SL, *Curr. Opin. Drug Discov. Devel.* 6(2): 237-52 (March 2003)) was conducted on test samples derived from tissues of nine human cancers and non-small-cell lung cancer (NSCLC) to identify EGFR mutations associated with these diseases.

TABLE 5 summarizes the results of DHPLC analysis of EGFR mutations in human colorectal cancer tissue.

TABLE 5
EGFR Mutations Identified in Human Colorectal Cancer

Exon	Mutation/SNP	Allelic Fraction	Unmutated Sequence	Mutated Sequence
Exon 4	AAC>AAT N158	heterozygous	CCCTGTGCAACGTG GAGAGCAT (SEQ ID NO:2)	CCCTGTGCAATGTG GAGAGCAT (SEQ ID NO:3)
Exon 4	AAC>AAT N158	homozygous	CCCTGTGCAACGTG GAGAGCAT (SEQ ID NO:2)	CCCTGTGCAATGTG GAGAGCAT (SEQ ID NO:3)
Exon 6	CAG>CGG Q217R	0.2	CATCTGTGCCCAGC AGTGCTCCGGGC (SEQ ID NO:4)	CATCTGTGCCCGGC AGTGCTCCGGGC (SEQ ID NO:5)
Exon 7	ins 105 bp R255-T290	0.15	Without insertion 105 bp	Insertion of 105 bp
Exon 12	A>T, 22bp 3' of exon	homozygous	TTTCTGTTTAGTTT ATGGAG (SEQ ID NO:6)	TTTCTGTTTTGTTT ATGGAG (SEQ ID NO:7)
Exon 12	A>T, 22bp 3' of exon	heterozygous	TTTCTGTTTAGTTT ATGGAG (SEQ ID NO:6)	TTTCTGTTTTGTTT ATGGAG (SEQ ID NO:7)
Exon 13	AGG>AAG, R521K	0.5	CCGGAGCCCAGGGA CTGCGTC (SEQ ID NO:8)	CCGGAGCCCAAGGA CTGCGTC (SEQ ID NO:9)
Exon 15	GCC>GCT, A613	heterozygous	GCAGACGCCGGCCA TGTG (SEQ ID NO:10)	GCAGACGCTGGCCA TGTG (SEQ ID NO:11)
Exon 16	ACT>ACA, T629	heterozygous	CAGATGCACTGGGC CAGGT (SEQ ID NO:12)	CAGATGCACAGGGC CAGGT (SEQ ID NO:13)
Exon 18	G>A 34bp 3' of exon	heterozygous	GGGCTGGGCCGCAG GGCCTCTC (SEQ ID NO:14)	GGGCTGGGCCACAG GGCCTCTC (SEQ ID NO:15)
Exon 18	G>A, 19bp 3' of exon	heterozygous	CCTGGCACAGGCCT CTGGGC (SEQ ID NO:16)	CCTGGCACAGACCT CTGGGC (SEQ ID NO:17)
Exon 20	CAG>CAA Q787	heterozygous	CACCGTGCAGCTCA TCACGC (SEQ ID NO:18)	CACCGTGCAACTCA TCACGC (SEQ ID NO:19)

[178] The EGFR genomic reference sequence was NT 079592. Four isoforms of human EGFR are reported in the LocusLink database. By default, the sequence of EGFR isoform a is referenced in the bioinformatics analysis. The mRNA and peptide reference sequences are

NM_005228 and NP_005219 respectively. Analysis of EGFR mutations in human lung cancer samples revealed the EGFR mutations, included, *e.g.*, G719S; L858R; E746_A750del; L747_E749del; and A750P. These EGFR mutations identified in NSCLC were different from the EGFR mutations identified in this invention. Bioinformatics analysis of selected EGFR mutations identified in this invention is detailed below in EXAMPLE 2.

EXAMPLE 2

BIOINFORMATICS ANALYSIS OF THE EGFR MUTATIONS OF THE INVENTION

[179] The EGFR missense mutations identified in human cancers (see TABLE 3 and TABLE 5) were analyzed using computational analysis tools to determine the effects of these mutations on EGFR function.

[180] *Comparison of Known EGFR Mutations and Coding SNPs with EGFR Missense Mutations.* The known EGFR mutations and single nucleotide polymorphisms (SNPs) in the coding region are summarized in TABLE 6 below. Sixteen (16) EGFR mutations have been reported. Nine (9) cSNPs of EGFR have been reported in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/index.html). Among them, eight (8) coding SNPs, N158N, P373P, R521K, Q787Q, T903T, D994D, T629T, and R836R, are also identified in the invention.

TABLE 6A
Known EGFR Non-synonymous SNPs and Mutations

Identifier	Variation type	cSNP Allele 1: frequency	Allele 2: frequency	Ref SNP
N158N	Synonymous	C	T	rs2072454
P373P	Synonymous	G: 0.993	A: 0.007	rs2302536
R521K	Non-synonymous	G	A	rs11543848
Q787Q	Synonymous	G	A	rs10251977
T903T	Synonymous	C: 0.925	T: 0.075	rs1140475
D994D	Synonymous	G:0.676	A:0.324	rs2293347
T629T	Synonymous	T: 0.685	A: 0.315	rs2072454
R836R	Synonymous	C: 0.75	T: 0.25	rs2229066
C977R	Synonymous	T	C	rs1140476

TABLE 6B

EGFR Mutation				
Identifier	Variation type	Wild type allele	Mutatant allele	Reference
G719S	Point mutation	T	G	[1]
L858R	Point mutation	T	G	[1]
delE746_A750	Deletion			[1,2]
delE746_T751insA	Deletion/insertion			[2]
delL747_E749	Deletion			[2]
delL747_E751insS	Deletion/insertion			[2]
delL747_P753insS	Deletion/insertion			[2]
L858M	Point mutation	C	T	[2]
G719C	Point mutation	G	T	[2]
A839T	Point mutation	G	A	[3]
K846R	Point mutation	A	G	[3]
delE747_T751				[3]
delE746_S752insD				[3]
V769M		G	A	[3]
L838V		C	G	[3]
E709A		A	C	[3]
E709G		A	G	[3]
L833V		T	G	[3]
H835L		A	T	[3]
L861Q		T	A	[3]
D761_E762insEAFQ				[3]
S768_D770dup				[3]
S768I		G	T	[3]
G719C		G	T	[3]
delL747_A750insP	Deletion/mutation			[3]
delE746_A750	Deletion			[3]
L861Q	Point mutation	T	A	[3]
W731Term	Point mutation	G	A	[3]
H773R	Point mutation	A	G	[3]
A859T	Point mutation			[4]
G719A	Point mutation			[4]
delE746_A750				[5]
delL747_T751				[5]
delL747_P753				[5]
delS752_I759				[5]
delL747_T751insP				[5]
delL757_P753insQ				[5]
delE746_A750insRP				[5]
delE746__T751insVA				[5]
delE746_S752insA				[5]
delL747_T751insQ				[5]
delL747_S752insQH				[5]
delL747_A750insP				[5]
delL747_P753insS				[5]
delT751_I759insS				[5]
E709H				[5]
T790M				[5]
S768I				[5]
R776C				[5]
V769L				[5]
I744_K745insKIPVAI				[5]
D761_E762insEAFQ				[5]
A767_S768insTLA				[5]
V769_D770insASV				[5]
D770_N771insY				[5]

[1] Science 2004; 304 1497-500
[2] J Clini Oncol. 2005, Feb 1. 857-865
[3] Clinical cancer research Vol 10, 8195-8203
[4] J Clin Oncol. 2005 Feb 14. [15710947]
[5] Cancer research 64, 8919-8923, Dec 15, 2004

[181] *Pfam Analysis of the Potential Effect of the EGFR Missense Mutations on EGFR Protein Domain Structure.* The effect of the EGFR missense mutations and non-synonymous polymorphisms on the protein domain structure of EGFR was analyzed using the Pfam computational analysis tool. Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein families based on the Swissprot 44.5 and SP-TrEMBL 27.5 protein sequence databases. A search using Pfam showed that the protein kinase domain of EGFR is located between amino acid positions 712 and 968 (score 263.9, E = 2.6e-76). The positions of the EGFR missense mutations identified in TABLE 3 which appear in the Pfam model of protein kinase domain are highlighted in bold underlined text. As shown in TABLE 5, twenty-eight (28) mutations identified in TABLE 3 are located in the protein kinase domain region and six (6) others are located in the receptor L domain. As shown in TABLE 7, alignment of the human wild- type EGFR sequence with the Pfam model of protein kinase domain indicates G729, K745, G779 and R932 are highly conserved, while G735, I740 and L760 are moderately conserved. Mutations that change amino acid residues at conserved positions may potentially alter the protein function.

TABLE 7
Sequence Alignment Comparison of Human EGFR with Pfam Model of Protein Kinase Domain

```
            *->lklgkkLGeGaFGeVykGtlkgsgegtkikVAVKtLkeigasseeig
               +k++k+LG+GaFG+VykG + ++ge+ ki+VA+K L+e  +s+++
EGFR   712      FKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKA-- 755

            redFlrEAsiMkklGdHpNiVrLlGvctkegePggpglyiVtEymegGdL
             ++ l EA +M +  d+p++ rLlG+c+          + ++t +m+ G+L
EGFR   756 NKEILDEAYVMASV-DNPHVCRLLGICL--TST----VQLITQLMPFGCL 798

            ldfLrkhregrpLtlkdLlsfalQiAkGMeYLesknfvHRDLAARNcLVs
            ld+ r+h  ++++  + Ll++++QiAkGM+YLe++++vHRDLAARN+LV
EGFR   799 LDYVREH--KDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVK 846

            enlvVKIsDFGLaRdi..ynddyYvrkkgggklPvkWmAPEslkygkFts
             + +VKI+DFGLa++++ ++++y      +ggk+P+kWmA+Es+ ++++t+
EGFR   847 TPQHVKITDFGLAKLLgaEEKEYH---AEGGKVPIKWMALESILHRIYTH 893

            kSDVWSFGVlLWEiftlGeqPFYpgmsneevlellyedGyRLprPenCPd
            +SDVWS+GV+ WE++t+G +P Y g++ +e+   l e+G+RLp+P+ C+
EGFR   894 QSDVWSYGVTVWELMTFGSKP-YDGIPASEISSIL-EKGERLPQPPICTI 941

            elYdlMlqCWaedPedRPtFselverL<-*
            ++Y++M +CW+ d++ RP F+el+ ++
EGFR   942 DVYMIMVKCWMIDADSRPKFRELIIEF    968
```

TABLE 8
EGFR Mutations Identified In Protein Function Domains

Domain	Seq-from	Seq-to	Mutations
Receptor L domain	57	168	
Furin-like cysteine rich region	184	338	Q217R, G221W, G239C, C251F, L267V, E282D
Receptor L domain	361	481	
Protein tyrosine kinase	712	968	T725P, Y727S, G729R, I732T, G735R, E736D, V738G, I740T, K745R, R748I, R748G, T751I, T751A, K754E, K754R, Y891S, N756S, K757R, K757E, E758G, L760P, N771D, V774M, G779S, I890T, Q894Term, Y900Term, R932G

[182] *Three-Dimensional Protein Modelling Analysis of the Effect of EGFR Mutation on the EGFR Protein Kinase Domain.* The effect of the EGFR missense mutations on the protein domain structure of EGFR was further analyzed using the three-dimensional structure of the EGFR protein kinase domain provided by the protein data bank (PDB). FIG. 1 is a schematic drawing of the three-dimensional structure of wild-type EGFR protein kinase domain rendered in Cn3d, a structure visualizing software provided by NCBI. Locations of EGFR missense mutation are highlighted by arrows. Position L760 is located in the middle of alpha-helix protein secondary structure. Proline can potentially break the helix structure. Change of leucine to proline at 760 (L760P) can have dramatic effect on the local secondary structure (See also *infra* Example 2, Section E).

[183] *NetPhos Analysis of the Effect of EGFR Missense Mutations on EGFR Protein Phosphorylation.* The EGFR is part of a subfamily of four closely related receptors EGFR (or ErbB-1), Her 2/neu (ErbB-2), Her 3 (ErbB-3) and Her 4 (ErbB-4). Receptors exist as inactive single units or monomers that, on activation by ligand binding, pair to form an active dimer. The two receptors that form a pair are not necessarily identical, for example an EGF-1 receptor (EGFR) may pair with another EGF-1 receptor, giving a so-called homodimer, or an EGFR may pair with another member of the receptor family, such as Her 2/*neu*, to give an asymmetrical heterodimer. Once pairing takes place, the tyrosine kinase enzyme in the intracellular domain of the receptor becomes activated, transphosphorylating both intracellular domains, and initiating the cascade of intracellular events which results in the signal reaching the nucleus. Wells A, *Int'l. J. Biochem. Cell Biol.*, 31:637-643 (1999).

[184] A recent publication from Sordella *et al.*, *Science* 305: 1163-1167 (2004) provided insight into the role of EGFR mutations in tumourogenesis. It has been demonstrated that ligand stimulation results in a different phosphorylation pattern in mutated receptors compared with that seen in wild-type receptors. As a result of this altered phosphorylation pattern, mutated receptors selectively activate cell survival pathways and thus cells with mutated receptors are relatively resistant to apoptosis.

[185] Accordingly, the effect of the EGFR missense mutations on known and potential protein phosphorylation sites of EGFR was analyzed. Known EGFR phosphorylation sites include Thr678, Thr693, Ser695, Tyr869, Ser1070, Ser1071, Tyr1092, Tyr1110, Tyr1172, and Tyr1197 which are summarized below in TABLE 9. The known EGFR phosphorylation sites are highlighted in bold text.

[186] Potential EGFR phosphorylation sites were identified by computational analysis using the NetPhos computational analysis tool. NetPhos produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins. Blom *et al.*, *J. Mol. Biol.* 294(5): 1351-1362 (1999). Potential EGFR phosphorylation sites predicted by NetPhos are summarized below in TABLE 9. The known EGFR phosphorylation sites are highlighted in bold text. Among these predicted phosphorylation sites identified, Y285 is close to E282, T693, also a known threonine phosphorylation site, is mutated in NSCLC, T725 is mutated in ovarian cancer, T751 is mutated in prostate cancer, S752 is next to T751 and close to other mutation sites K754 and N756, Y764 is close to L760, T892 is adjacent to Y891. In conclusion, mutations at the positions E282, T693, T725, T751, K754, N756, L760, and Y891 may influence the phosphorylation patterns of nearby sites.

TABLE 9
EGFR Phosphorylation Sites Predicted by NetPhos

Phosphorylation	Amino Acid Position
Serine	123, 227, 229, 246, 286, 315, 380, 452, 457, 484, 498, 530,720, *752*, 957, 991, 1028, 1030, 1037, 1042, 1070, 1071, 1081, 1104, 1149,1166, 1190, 204
Threonine	259, 397, 572, **678**, ***693***, *725*, *751*, *892*, 940, 1029, 1131
Tyrosine	69, 88,113,117, *285*, 316, 471, 585, 610, 626, *764*, *901*, **869**, 1016, **1092**, **1110**,1125

[187] The change in phosphorylation pattern by mutation is investigated by NetPhos, which is summarized in TABLE 10. For example, I732T, N756S, N771D and Y891S each may introduce new phosphorylation sites, and Y891S may also increase the phosphorylation potential of T892.

TABLE 10
Effect of Mutation on Phosphorylation.

	WT	C251F	E282D	D587N	G729R	I732T	R748G	R748I	T751I	T751A	K754E	K754R	N756S	L760P	N771D	I890T	Y891S	M952R
S246	0.724	0.747																
S286	0.698		0.668															
S752	0.997						0.994	0.995	0.995	0.996	0.996	0.994						
S756													0.992					
S768															0.976			
S891																	0.994	
S957	0.825																	0.934
T725	0.547				0.62													
T732						0.671												
T751	0.877						0.51				0.967	0.85						
T892	0.535															0.536	0.739	
Y285	0.951		0.859															
Y585	0.826			0.789														
Y764	0.881													0.972				

[188] A score highlighted in red colour indicates the corresponding mutation increases phosphorylation potential of a nearby residue, a score in green colour indicates the mutation decreases the phosphorylation potential of a nearby residue.

[189] *PROSITE Analysis of the Potential Effect of EGFR Missense Mutations on Other EGFR Protein Regulatory Sites.* The effect of the EGFR missense mutations on other protein regulatory sites was analyzed using the PROSITE computational analysis tool. PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs as well as to identify potential sites for protein modification. Hulo N *et al.*, *Nucl. Acids. Res.* 32:D134-D137 (2004); Sigrist CJA *et al.*, *Brief Bioinform.* 3:265-274(2002); Gattiker A *et al.*, *Applied Bioinformatics* 1:107-108(2002). A search using PROSITE showed that the other potential sites for protein modification are predicted at EGFR amino acid positions, 220-222 and 752-754 as PKC phosphorylation sites as summarized below in TABLE 11. Also, EGFR amino acids 718-745 matched to consensus of an ATP-binding site. Eight (8) mis-sense mutations T725P, Y727S, G729S, I732T, G735R, E736D, V738G and I740T are located within the ATP-binding site. These mutations may potentially influence the ATP binding.

TABLE 11
Potential EGFR protein modification sites predicted by PROSITE

Function	Positions
ATP-binding region	**718-745**
Tyrosine kinase active site	833-845
Cysteine-rich	187-264
N-myristoylation	5-10, 197-202, 322-327, 339-344, 635-640, 649-654, 721-726, 779-784, 917-922, 1185-1190
PKC phosphorylation	35-37, 220-222, 259-261, 397-399, 452-454, 492-494, 498-500, 525-527, 572-574, 678-680, **752-754**, 1029-1031
CK2 phosphorylation	43-46, 81-84, 123-126, 227-230, 259-262, 397-400, 452-455, 457-460, 903-906, 925-928, 1081-1084, 1149-1152, 1190-1193
Tyrosine phosphorylation	128-131, 175-178, 196-198, 352-355, 361-364, 413-416, 444-447, 528-531, 568-571, 603-606, 623-626, 1043-1046, 1044-1047, 1094-1097, 1148-1151
Lysine-rich region	439-481
Cell attachment	377-379
cAMP/cGMP dependent phosphorylation	675-678
Tyrosine sulphation	309-323, 1009-1023, 1165-1179

[190] *ClustalW Polypeptide Alignment and Sequence Analysis to Estimate the Potential Effect of EGFR Missense Mutations on EGFR Function.* The effect of EGFR missense mutations on EGFR biological function is further analyzed by peptide sequence alignment. Known EGFR sequences of various organisms including cow, chimpanzee, chicken, human, mouse, rat, pig, zebrafish, dog, the CG10079-PA polypeptide of fruit fly, and the CG10079-PB polypeptide of fruit fly were obtained from GenBank and aligned using ClustalW. Chenna *et al.*, *Nucleic Acids Res.*, 31 (13):3497-500 (2003). Clustal W is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

[191] For every position with a mutation reported, the mutated residues were inspected for their occurrence in organisms other than human. It is hypothesized that if the mutated residue is present in the wild-type sequence of another species in the corresponding position, the amino acid change may not have any adverse effect on the protein function. The results of ClustalW analysis are summarized below in TABLE 12. In position 756 of human EGFR, a mutation of asparagine to serine was found present in fruit fly, a non-human species.

TABLE 12

Summary of sequence alignment of wild type EGFR sequences from multiple organisms

Mutation	Variation of amino acids
Q217R	"E" is present in zebrafish; "D" is present in fruit fly
G221W	"H" is present in mouse; "R" is present in rat
G239C	No variation is observed at this position
C251F	No variation is observed at this position
L267V	"M" is found in fruit fly
E282D	"L" is present in pig; "Y" is present in zebrafish
L509V	"M" is present in zebrafish; "Q" is present in fruit fly
R521K	"F" is present in fruit fly
D587N	No variation is observed at this position
C595W	No variation is observed at this position
H618Y	"Q" is present in chimp and zebrafish; "F" is present in fruit fly
T693I	"I" is present in chimp, "R" is present in fruit fly
P699A	"A" is present in fruit fly
E709V	"D" is present in fruit fly
T725P	"R" is present in fruit fly
Y727S	"H" is present in zebrafish
G729S	No variation is observed at this position
I732T	"V" is present in zebrafish and fruit fly
G735R	No variation is observed at this position
V738G	No variation is observed at this position
I740T	No variation is observed at this position
K745R	No variation is observed at this position
R748G	"L" is present in fruit fly
R748I	"L" is present in fruit fly
T751I	No variation is observed at this position
T751A	No variation is observed at this position
K754E	"E" is present in fruit fly
K754R	"E" is present in fruit fly
N756S	"S" is present in fruit fly
K757E	"E" is present in fruit fly
K757R	"E" is present in fruit fly
E758G	No variation is observed at this position
L760P	"M" is present in zebrafish
N771D	"H" is present in zebrafish and fruit fly
V774M	"L" is present in fruit fly
G779S	"A" is present in fruit fly
I890T	"T" is present in zebrafish, "V" is present in pig and fruit fly
Y891S	"F" is present in fruit fly
Q894Term	"K" is present in fruit fly
Y900Term	"F" is present in fruit fly
R932G	"K" is present in fruit fly
M971R	"H" is present in fruit fly

[192] *The Effect of EGFR Missense Mutation on Amino Acid Property.* The change of amino acid property observed by EGFR mutation (Valdar WS, *Proteins* 48(2): 227-41 (2002)) is summarized in TABLE 13.

TABLE 13
Influence of EGFR mutations on EGFR protein secondary structure

Mutation	Amino acid property change
Q217R	Polar->positive
G221W	Tiny->aromatic
G239C	
C251F	Tiny->aromatic
E282D	
L509V	
R521K	
D587N	Negative->polar
C595W	Tiny->aromatic
H618Y	Positive -> aromatic
T693I	Small->Aliphatic
P699A	Proline->tiny
E709V	Negative->Aliphatic
T725P	Small->praline
Y727S	Aromatic->tiny
G729S	
I732T	Aliphatic->small
G735R	Tiny->positive
E736D	
V738G	Aliphatic->tiny
I740T	Aliphatic->small
K745R	
R748G	Positive->tiny
R748I	Positive->aliphatic
T751I	Small->aliphatic
T751A	Small->tiny
K754E	Positive->negative
K754R	
N756S	Small->tiny
Y891S	Aromatic->tiny
K757E	Positive->negative
K757R	
E758G	Negative->tiny
L760P	Aliphatic->praline
N771D	Polar->negative
V774M	
G779S	
I890T	Aliphatic->small
Y891S	Aromatic->tiny
Q894Term	
Y900Term	
R932G	Positive->tiny
M971R	Hydrophobic->positive

[193] Q217R, G725R and M971R each introduces a positive charge, N771D introduces a negative charge, H618Y, R748G, R748I and R932G each loses a positive charge, D587N,

E709V and E758G each loses a negative charge, K754E and K757E each converts a positive charge to negative one.

[194] *nnPredict Analysis of the Effect of EGFR Missense Mutations on EGFR Secondary Structure*. Secondary structure prediction of wild-type (TABLE 14 and TABLE 15) and mutated EGFR sequences (see TABLES 12-27) were performed by nnPredict. Kneller DG, Cohen FE & Langridge R, *J. Mol. Biol.* (214) 171-182 (1990). All TABLES (*e.g.*, TABLES 11, 13, 15, 17, 19, 21, 23, 25 and 27) in this section that summarize the EGFR protein secondary structure as predicted by nnPredict use "H", "E" and a dash "-" as identifiers, which are defined as follows. An "H" designates helix protein secondary structure. An "E" designates strand protein secondary structure. A dash, *i.e.*, "-", designates no prediction of protein secondary structure. As appropriate, the position of the mutated amino acid residue is highlighted as with a grey shaded box.

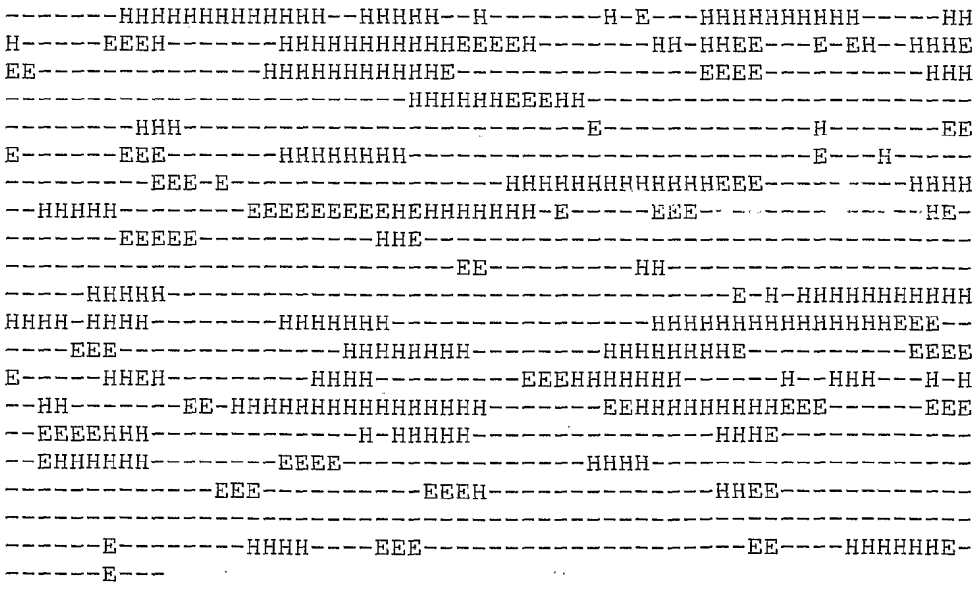
[195] The amino acid sequence of wild-type EGFR polypeptide (SEQ ID NO:1) is shown below in TABLE 14.

TABLE 14

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:1)

[196] A schematic representation of the secondary structure of wild-type EGFR polypeptide (SEQ ID NO:1) predicted using nnPredict analysis is shown below in TABLE 15.

TABLE 15



[197] The amino acid sequence of EGFR mutant polypeptide Q217R (SEQ ID NO:36) is shown below in TABLE 16. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 16

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICA**r**QCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:36)

[198] A schematic representation of the secondary structure of EGFR mutant polypeptide Q217R (SEQ ID NO:36) predicted using nnPredict analysis is shown below in TABLE 17.

TABLE 17

-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHHE---------------EEEE----------HHH
------------------------HHHHHHEEEE-H----------------------
--------HHH-------------------------E-------------H-------EE
E------EEE-------HHHHHHHH-----------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE----------HHE-----------------------------------
----------------------------EE---------HH------------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH--------HHHHHHH----------------HHHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE-----------
--EHHHHHH--------EEEE---------------HHHH-----------------
-------------EEE----------EEEH-------------HHEE-----------

------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---

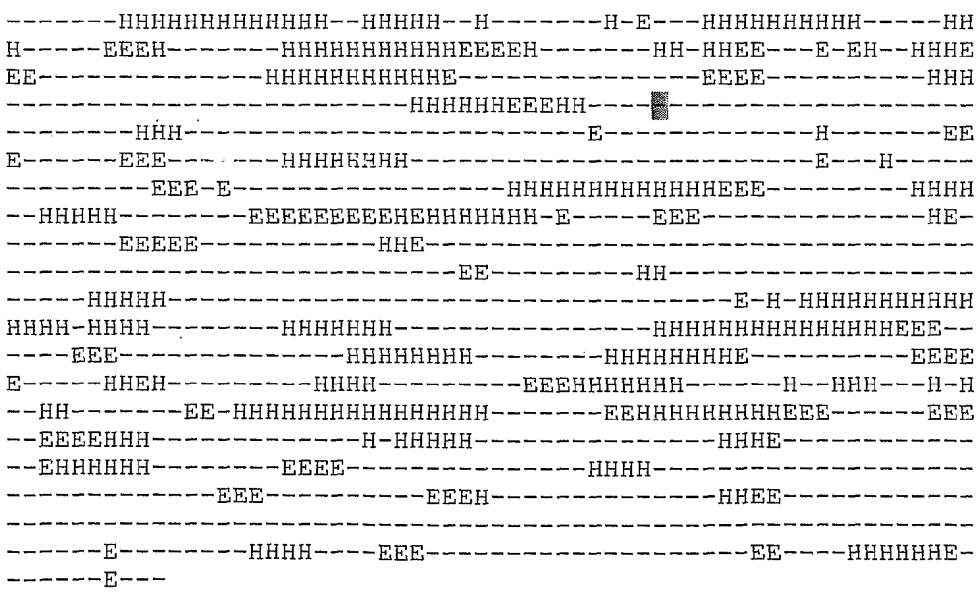
[199] The amino acid sequence of EGFR mutant polypeptide R521K (SEQ ID NO:37) is shown below in TABLE 18. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 18

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCS**W**RCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:37)

[200] A schematic representation of the secondary structure of EGFR mutant polypeptide G221W (SEQ ID NO:37) predicted using nnPredict analysis is shown below in TABLE 19.

TABLE 19



[201] The amino acid sequence of EGFR mutant polypeptide G239C (SEQ ID NO:38) is shown below in TABLE 20. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 20

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAA**c**C
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:38)

[202] A schematic representation of the secondary structure of EGFR mutant polypeptide G239C (SEQ ID NO:38) predicted using nnPredict analysis is shown below in TABLE 21.

TABLE 21

```
-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHHE--------------EEEE----------HHH
------------------------HHHHHHEEEHH--------------------▒-
--------HHH-------------------------E-------------H--------EE
E-----  -EEE-------HHHHHHHH-------------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE-----------HHE---------------------------------
----------------------------EE---------HH-------------------
-----HHHHH-------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH--------HHHHHHH---------------HHHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE-----------
--EHHHHHH--------EEEE---------------HHHH-------------------
------------EEE----------EEEH-------------HHEE-----------
-----------------------------------------------------------
------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---
```

[203] The amino acid sequence of EGFR mutant polypeptide C251F (SEQ ID NO:39) is shown below in TABLE 22. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 22

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLV**F**RKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:39)

[204] A schematic representation of the secondary structure of EGFR mutant polypeptide C251F (SEQ ID NO:39) predicted using nnPredict analysis is shown below in TABLE 23.

TABLE 23

```
-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHHE--------------EEEE----------HHH
-------------------------HHHHHHEEEHH-------------------------
-------HHHHH----------------------E-------------H--------EE
E------EEE-------HHHHHHHH-----------------------E ---------
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE-----------HHE---------------------------------
----------------------------EE---------HH-------------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH--------HHHHHHH----------------HHHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE-----------
--EHHHHHH--------EEEE--------------HHHH-------------------
-------------EEE----------EEEH-------------HHEE-----------
-------------------------------------------------------------
------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---
```

[205] The amino acid sequence of EGFR mutant polypeptide L267V (SEQ ID NO:40) is shown below in TABLE 24. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 24

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPP**v**MLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:40)

[206] A schematic representation of the secondary structure of EGFR mutant polypeptide L267V (SEQ ID NO:40) predicted using nnPredict analysis is shown below in TABLE 25.

TABLE 25

-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHHE--------------EEEE----------HHH
------------------------HHHHHHEEEHH----------------------
--------HHH-------------H---------E-------------H-------EE
E------EEE-------HHHHHHHH-------------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE----------HHE-----------------------------------
----------------------------EE---------HH------------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHH
HHHH-HHHH--------HHHHHHH---------------HHHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHH-------EEHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE----------
--EHHHHHH--------EEEE--------------HHHH-----------------
-------------EEE----------EEEH-------------HHEE----------

------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---

[207] The amino acid sequence of EGFR mutant polypeptide E282D (SEQ ID NO:41) is shown below in TABLE 26. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 26

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNP**D**GKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO: 41)

[208] A schematic representation of the secondary structure of EGFR mutant polypeptide E282D (SEQ ID NO:41) predicted using nnPredict analysis is shown below in TABLE 27.

TABLE 27

```
-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHE---------------EEEE----------HHH
------------------------HHHHHHEEEHH-----------------------
--------HHH-------------------------E----H--------H-------EE
E------EEE-------HHHHHHH-----------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHH-E-----EEE--------------HE-
-------EEEEE-----------HHE---------------------------------
----------------------------EE---------HH-------------------
-----HHHHH-------------------------------E-H-HHHHHHHHHH
HHHH-HHHH--------HHHHHHH----------------HHHHHHHHHHHHHHEEE--
----EEE--------------HHHHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE-----------
--EHHHHHH--------EEEE--------------HHHH-------------------
-------------EEE----------EEEH-------------HHEE-----------
-----------------------------------------------------------
------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---
```

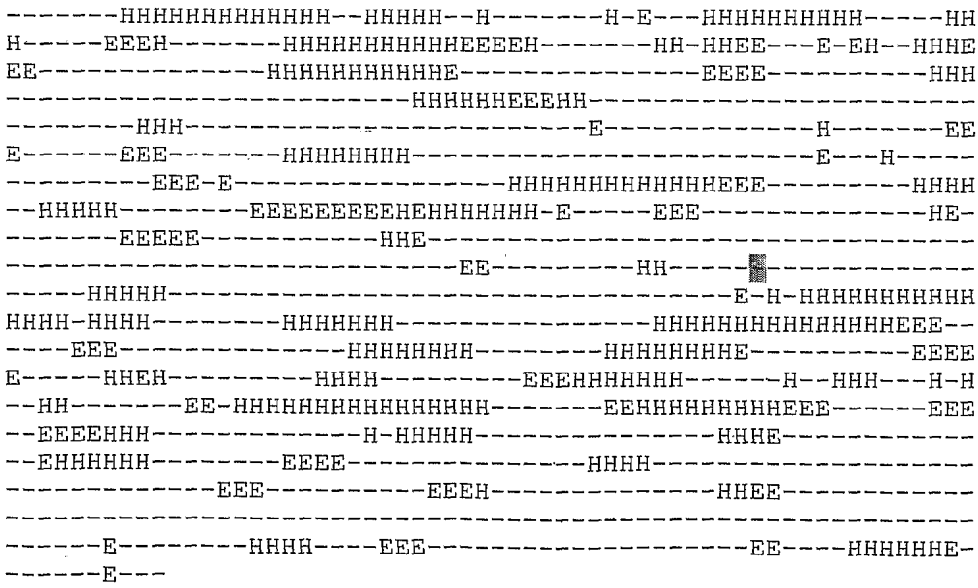
[209] The amino acid sequence of EGFR mutant polypeptide D587N (SEQ ID NO:42) is shown below in TABLE 28. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 28

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYI**N**GPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO: 42)

[210] A schematic representation of the secondary structure of EGFR mutant polypeptide D587N (SEQ ID NO:42) predicted using nnPredict analysis is shown below in TABLE 29.

TABLE 29



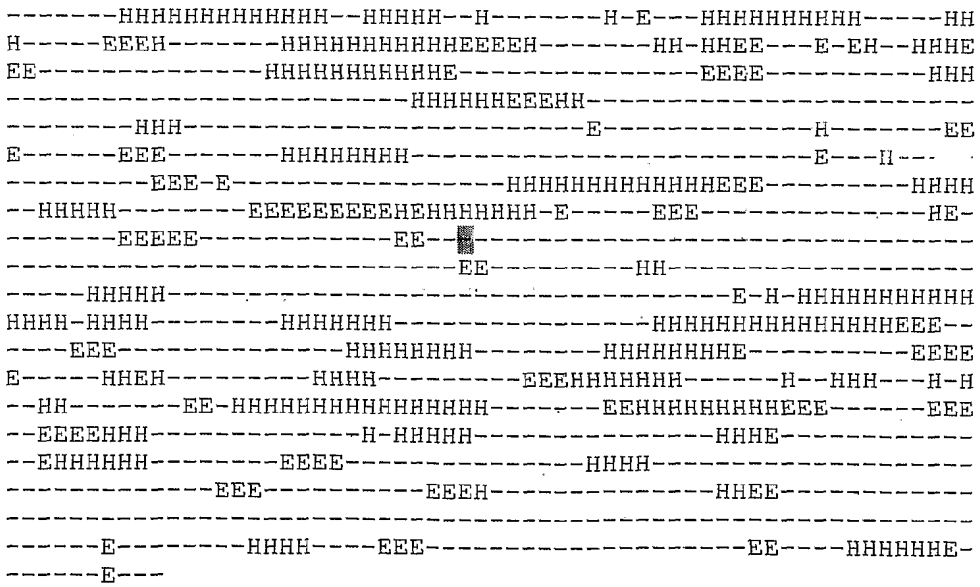
[211] The amino acid sequence of EGFR mutant polypeptide L509V (SEQ ID NO:43) is shown below in TABLE 30. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 30

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHA**v**CSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:43)

[212] A schematic representation of the secondary structure of EGFR mutant polypeptide L509V (SEQ ID NO:43) predicted using nnPredict analysis is shown below in TABLE 31.

TABLE 31



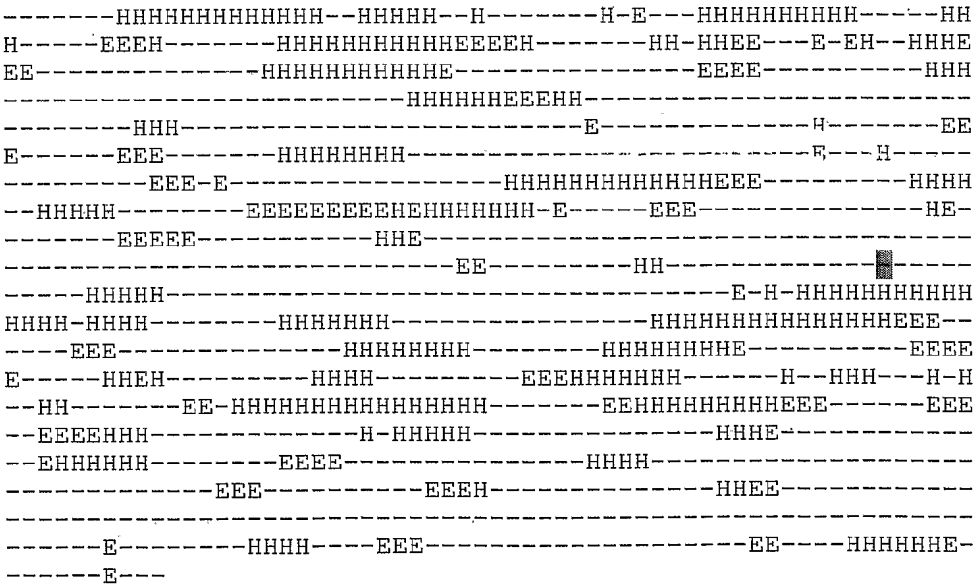
[213] The amino acid sequence of EGFR mutant polypeptide C595W (SEQ ID NO:44) is shown below in TABLE 32. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 32

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKT**W**PAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:44)

[214] A schematic representation of the secondary structure of EGFR mutant polypeptide C595W (SEQ ID NO:44) predicted using nnPredict analysis is shown below in TABLE 33.

TABLE 33



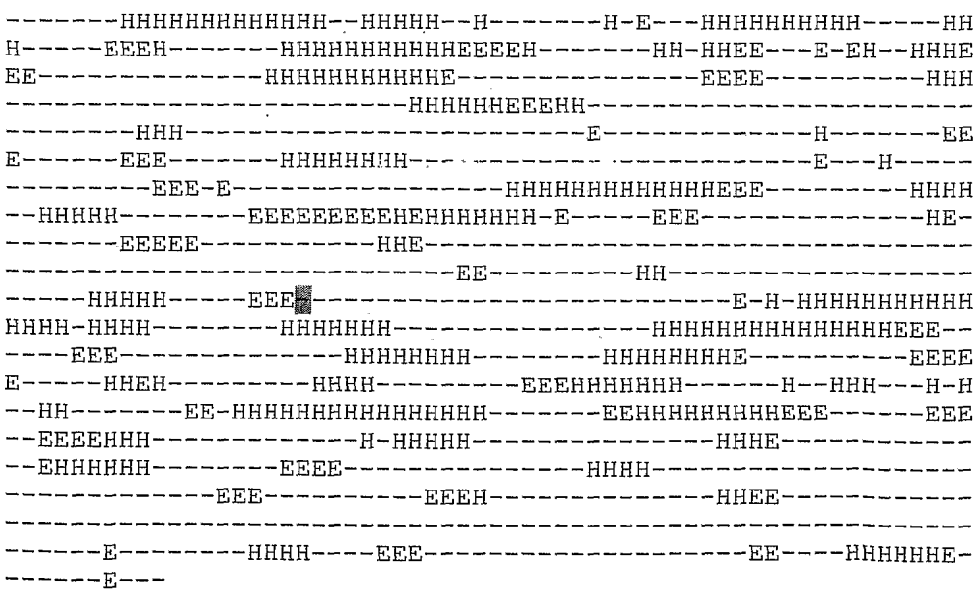
[215] The amino acid sequence of EGFR mutant polypeptide H618Y (SEQ ID NO:45) is shown below in TABLE 34. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 34

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVC**Y**LCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:45)

[216] A schematic representation of the secondary structure of EGFR mutant polypeptide H618Y (SEQ ID NO:45) predicted using nnPredict analysis is shown below in TABLE 35.

TABLE 35



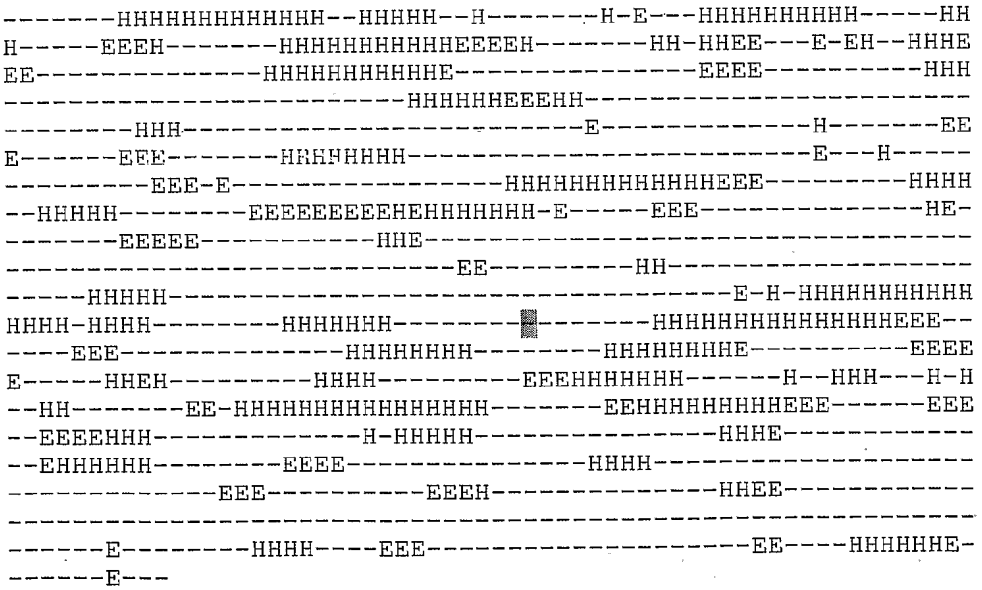
[217] The amino acid sequence of EGFR mutant polypeptide T693I (SEQ ID NO:46) is shown below in TABLE 36. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 36

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPL**I**PSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:46)

[218] A schematic representation of the secondary structure of EGFR mutant polypeptide T693I (SEQ ID NO:46) predicted using nnPredict analysis is shown below in TABLE 37.

TABLE 37



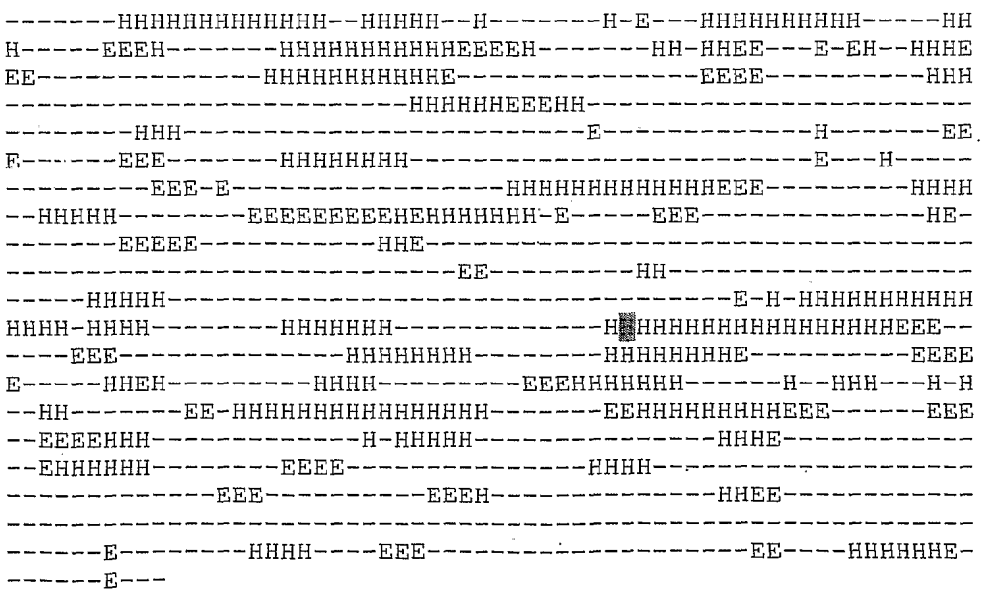
[219] The amino acid sequence of EGFR mutant polypeptide P699A (SEQ ID NO:47) is shown below in TABLE 38. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 38

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEA**A**NQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:47)

[220] A schematic representation of the secondary structure of EGFR mutant polypeptide P699A (SEQ ID NO:47) predicted using nnPredict analysis is shown below in TABLE 39.

TABLE 39



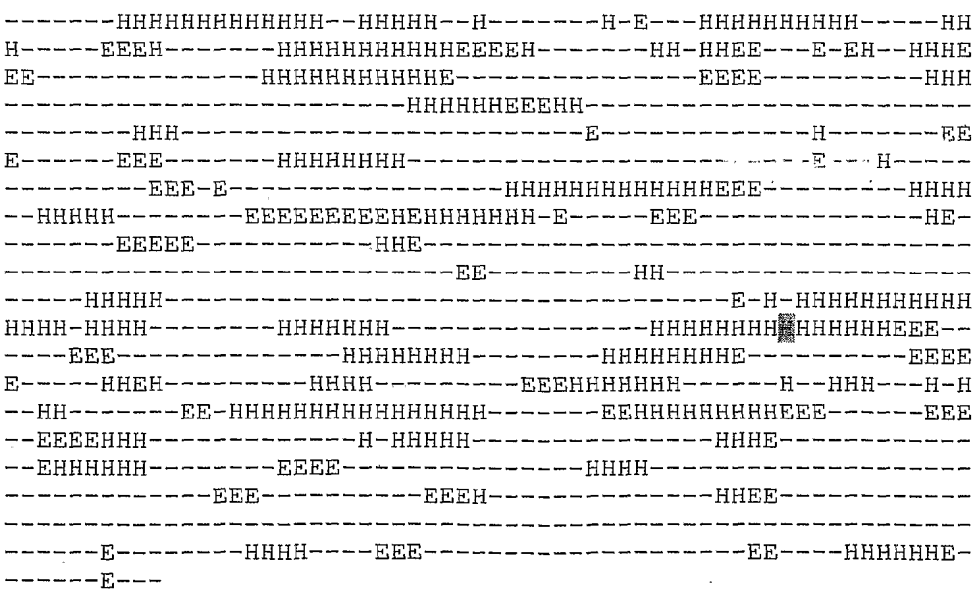
[221] The amino acid sequence of EGFR mutant polypeptide E709V (SEQ ID NO:48) is shown below in TABLE 40. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 40

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILK**v**TEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:48)

[222] A schematic representation of the secondary structure of EGFR mutant polypeptide E709V (SEQ ID NO:48) predicted using nnPredict analysis is shown below in TABLE 41.

TABLE 41



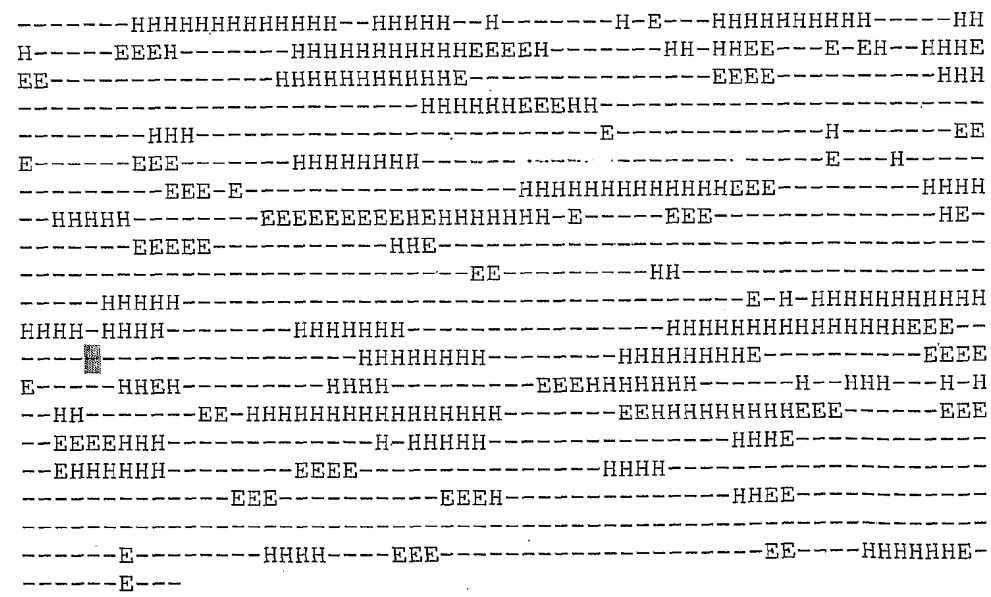
[223] The amino acid sequence of EGFR mutant polypeptide T725P (SEQ ID NO:49) is shown below in TABLE 42. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 42

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGPVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:49)

[224] A schematic representation of the secondary structure of EGFR mutant polypeptide T725P (SEQ ID NO:49) predicted using nnPredict analysis is shown below in TABLE 43.

TABLE 43



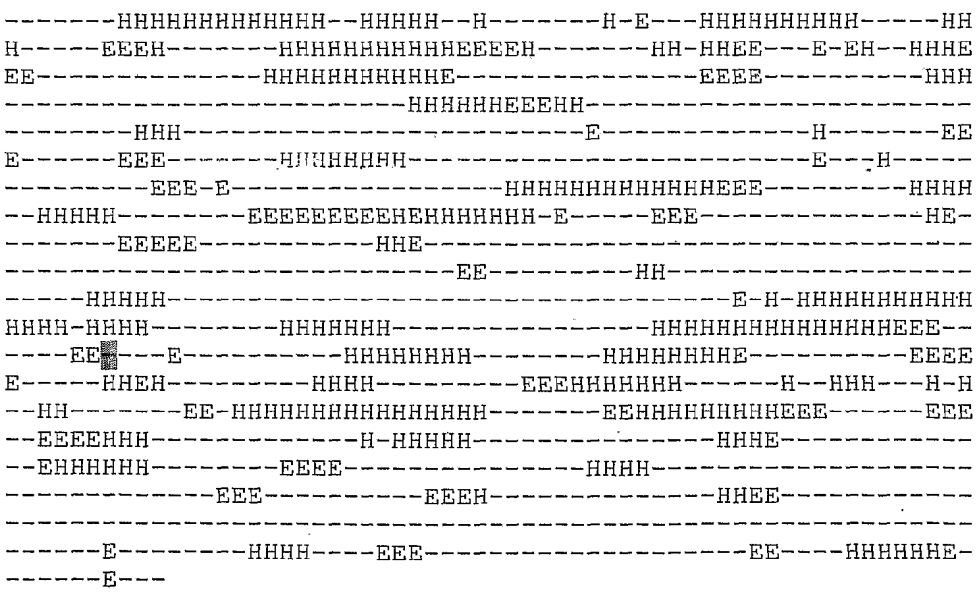
[225] The amino acid sequence of EGFR mutant polypeptide Y727S (SEQ ID NO:50) is shown below in TABLE 44. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 44

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTV**s**KGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:50)

[226] A schematic representation of the secondary structure of EGFR mutant polypeptide Y727S (SEQ ID NO:50) predicted using nnPredict analysis is shown below in TABLE 45.

TABLE 45



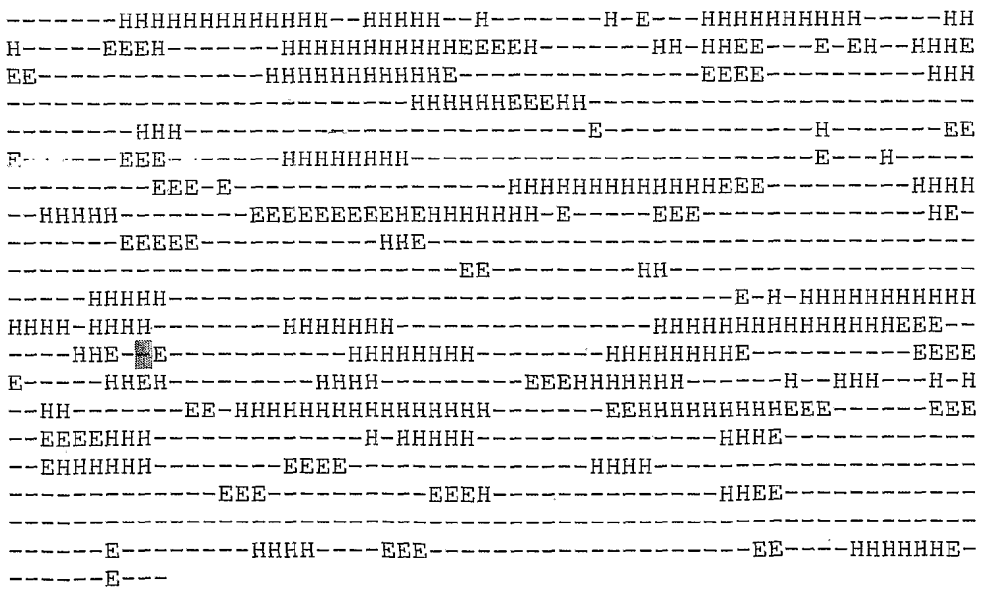
[227] The amino acid sequence of EGFR mutant polypeptide G729R (SEQ ID NO:51) is shown below in TABLE 46. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 46

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYK**R**LWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:51)

[228] A schematic representation of the secondary structure of EGFR mutant polypeptide G729R (SEQ ID NO:51) predicted using nnPredict analysis is shown below in TABLE 47.

TABLE 47



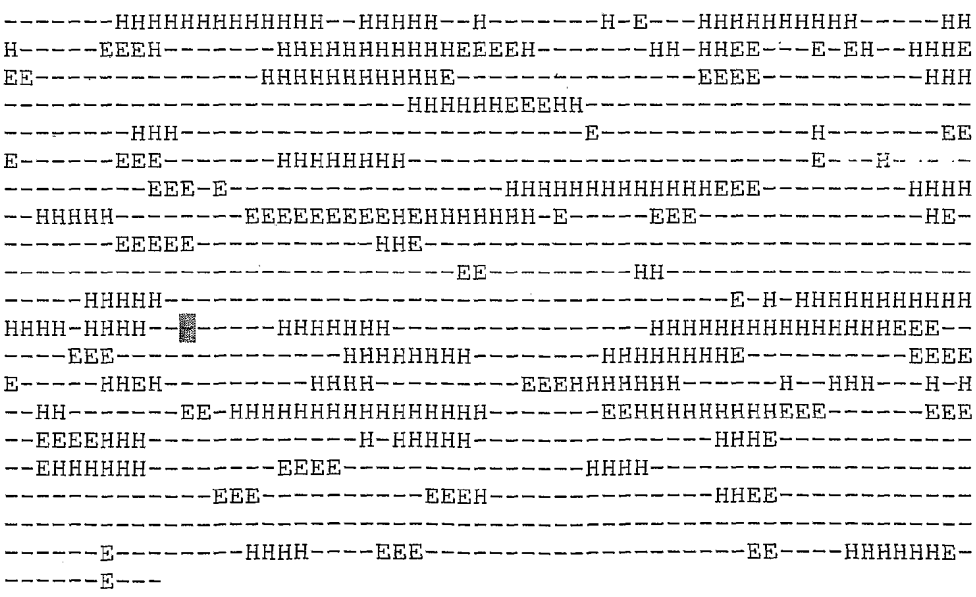
[229] The amino acid sequence of EGFR mutant polypeptide I732T (SEQ ID NO:52) is shown below in TABLE 48. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 48

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLW**T**PEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:52)

[230] A schematic representation of the secondary structure of EGFR mutant polypeptide I732T (SEQ ID NO:52) predicted using nnPredict analysis is shown below in TABLE 49.

TABLE 49



[231] The amino acid sequence of EGFR mutant polypeptide G735R (SEQ ID NO:53) is shown below in TABLE 50. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 50

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPE**R**EKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:53)

[232] A schematic representation of the secondary structure of EGFR mutant polypeptide G735R (SEQ ID NO:53) predicted using nnPredict analysis is shown below in TABLE 51.

TABLE 51

```
-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHE---------------EEEE----------HHH
------------------------HHHHHHEEEHH-----------------------
--------HHH-------------------------E-------------H-------EE
E------EEE-------HHHHHHHH-----------------------E--H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE-----HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE----------HHE---------------------------------
-----------------------EE---------HH-------------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHH
HHHH-HHHH--------HHHHHHH---------------HHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHH--------HHHHHHHE----------EEEE
E-----HHEH----H----HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHH-------EEHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------HHHE-----------
--EHHHHHH--------EEEE--------------HHHH-------------------
------------EEE----------EEEH-------------HHEE-----------
-----------------------------------------------------------
------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---
```

[233] The amino acid sequence of EGFR mutant polypeptide E736D (SEQ ID NO:54) is shown below in TABLE 52. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 52

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEG**D**KVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:54)

[234] A schematic representation of the secondary structure of EGFR mutant polypeptide E736D (SEQ ID NO:54) predicted using nnPredict analysis is shown below in TABLE 53.

TABLE 53

-------HHHHHHHHHHHHH--HHHHH--H--------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE--------------HHHHHHHHHHE---------------EEEE----------HHH
------------------------HHHHHHEEEHH-----------------------
--------HHH-------------------------E-------------H-------EE
E------EEE-------HHHHHHHH-----------------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE-----------HHE---------------------------------
-----------------------------EE---------HH-----------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH--------HHHHHHH-----------HHHHHHHHHHHHHHHEEE--
----EEE--------H-----EHHHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH--------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH-------------HHHE-----------
--EHHHHHH--------EEEE-------------HHHH-----------------
-------------EEE----------EEEH-------------HHEE----------

------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---

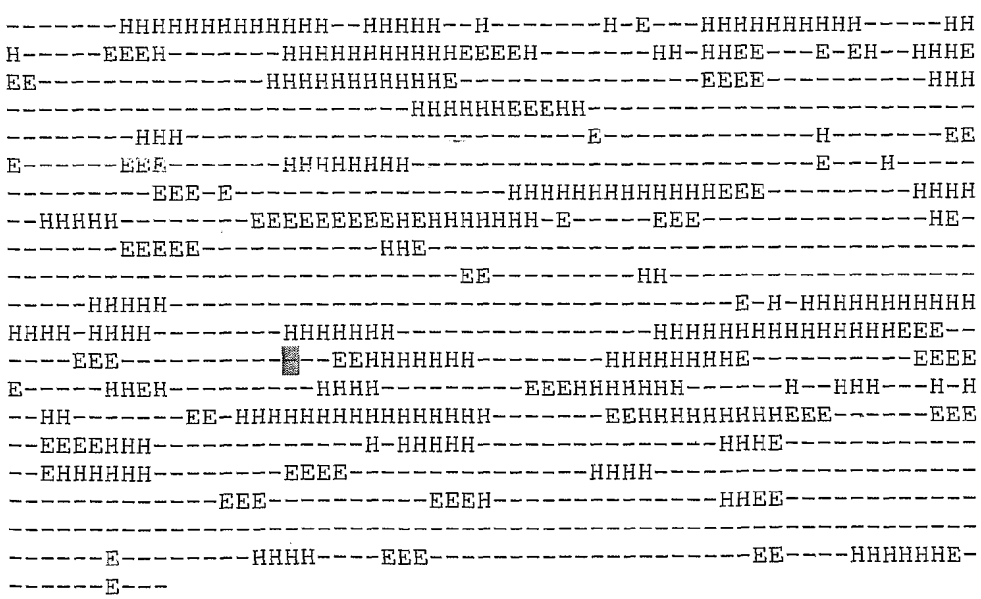
[235] The amino acid sequence of EGFR mutant polypeptide V738G (SEQ ID NO:55) is shown below in TABLE 54. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 54

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEK**G**KIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:55)

[236] A schematic representation of the secondary structure of EGFR mutant polypeptide V738G (SEQ ID NO:55) predicted using nnPredict analysis is shown below in TABLE 55.

TABLE 55



[237] The amino acid sequence of EGFR mutant polypeptide I740T (SEQ ID NO:56) is shown below in TABLE 56. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 56

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVK**T**PVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:56)

[238] A schematic representation of the secondary structure of EGFR mutant polypeptide I740T (SEQ ID NO:56) predicted using nnPredict analysis is shown below in TABLE 57.

TABLE 57

```
-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHHE--------------EEEE----------HHH
-------------------------HHHHHHEEEHH----------------------
-L------HHH-------------------------E-------------H-------EE
E------EEE-------HHHHHHHH-------------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE--------------HE-
--------EEEEE----------HHE-------------------------------
---------------------------EE---------HH-------------------
-----HHHHH-------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH--------HHHHHH---------------HHHHHHHHHHHHHHHEEE--
----EEE----------------HHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE-----------
--EHHHHHH--------EEEE---------------HHHH--------------------
-------------EEE----------EEEH-------------HHEE------------
-----------------------------------------------------------
------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---
```

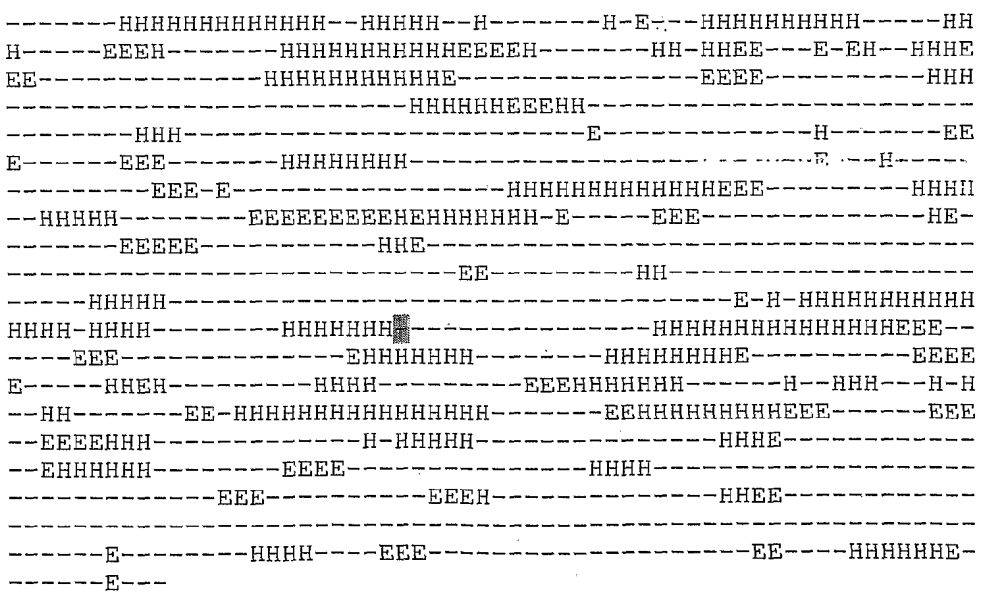
[239] The amino acid sequence of EGFR mutant polypeptide K745R (SEQ ID NO:57) is shown below in TABLE 58. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 58

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAI**R**ELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:57)

[240] A schematic representation of the secondary structure of EGFR mutant polypeptide K745R (SEQ ID NO:57) predicted using nnPredict analysis is shown below in TABLE 59.

TABLE 59



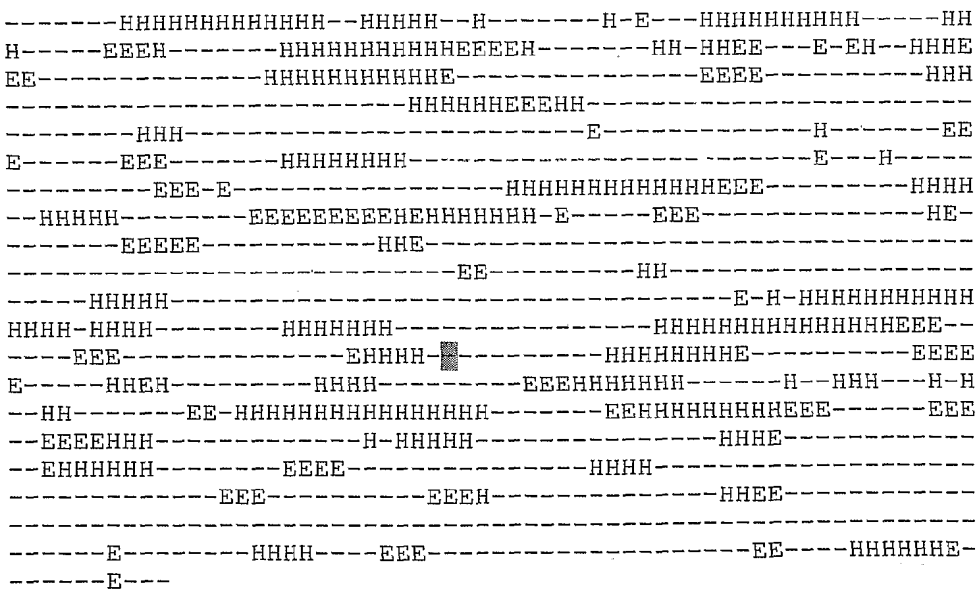
[241] The amino acid sequence of EGFR mutant polypeptide R748G (SEQ ID NO:58) is shown below in TABLE 60. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 60

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELGEATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:58)

[242] A schematic representation of the secondary structure of EGFR mutant polypeptide R748G (SEQ ID NO:58) predicted using nnPredict analysis is shown below in TABLE 61.

TABLE 61



[243] The amino acid sequence of EGFR mutant polypeptide R748I (SEQ ID NO:59) is shown below in TABLE 62. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 62

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKITLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKEL**I**EATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:59)

[244] A schematic representation of the secondary structure of EGFR mutant polypeptide R748I (SEQ ID NO:59) predicted using nnPredict analysis is shown below in TABLE 63.

TABLE 63

```
-------HHHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE-------------HHHHHHHHHHHE--------------EEEE----------HHH
------------------------HHHHHHEEEHH----------------------
--------HHH-------------------------E-------------H-------EE
E------EEE-------HHHHHHHH----------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE----------HHE---------------------------------
-----------------------EE---------HH------------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH-H-----HH---HHH---------------HHHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHHH-------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE----------
--EHHHHHH--------EEEE---------------HHHH-----------------
------------EEE----------EEEH-------------HHEE-----------
-----------------------------------------------------------
------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---
```

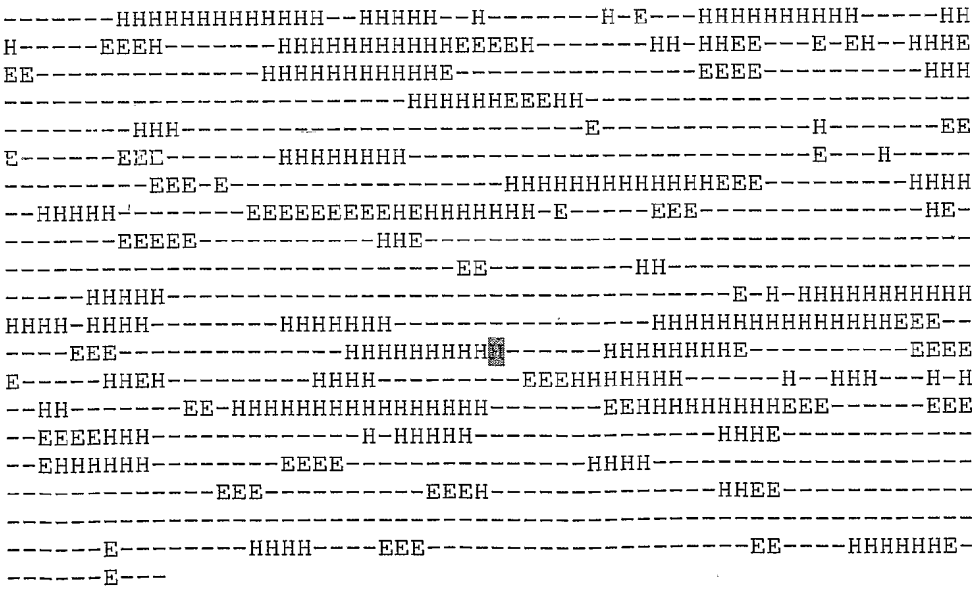
[245] The amino acid sequence of EGFR mutant polypeptide T751I (SEQ ID NO:60) is shown below in TABLE 64. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 64

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREAISPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:60)

[246] A schematic representation of the secondary structure of EGFR mutant polypeptide T751I (SEQ ID NO:60) predicted using nnPredict analysis is shown below in TABLE 65.

TABLE 65



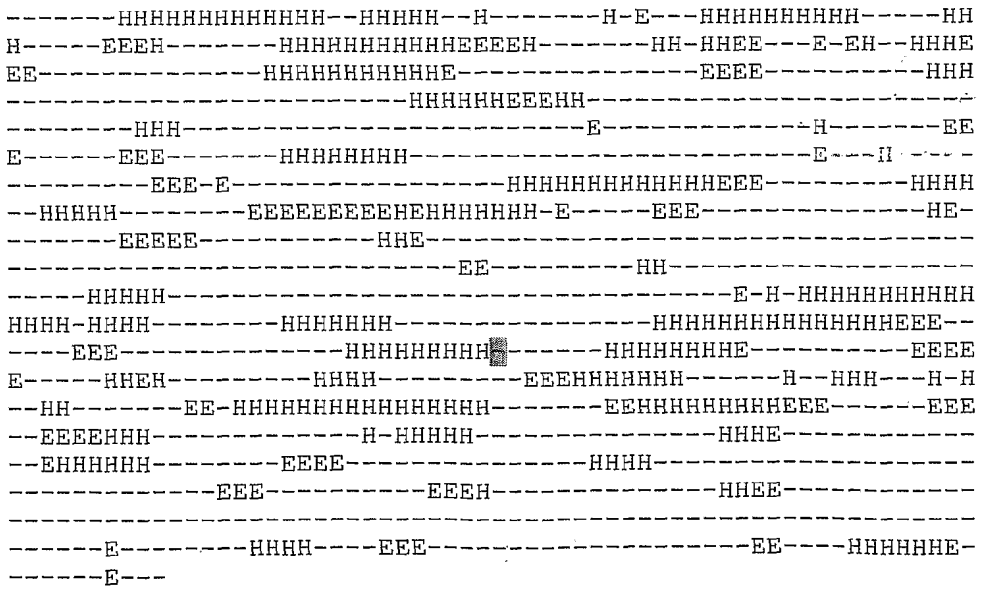
[247] The amino acid sequence of EGFR mutant polypeptide T751A (SEQ ID NO:61) is shown below in TABLE 66. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 66

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREAASPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:61)

[248] A schematic representation of the secondary structure of EGFR mutant polypeptide T751A (SEQ ID NO:61) predicted using nnPredict analysis is shown below in TABLE 67.

TABLE 67



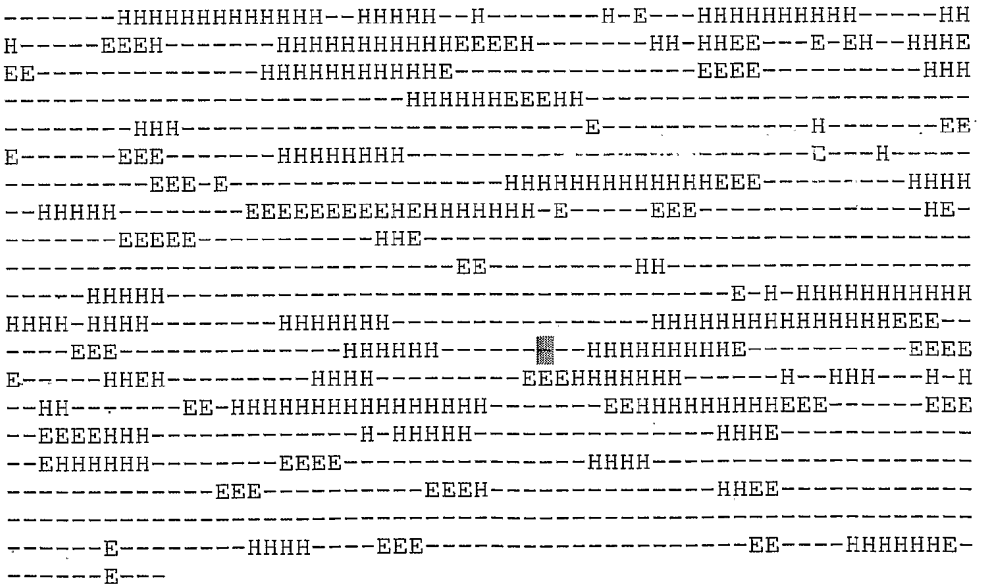
[249] The amino acid sequence of EGFR mutant polypeptide K754E (SEQ ID NO:62) is shown below in TABLE 68. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 68

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSP**E**ANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:62)

[250] A schematic representation of the secondary structure of EGFR mutant polypeptide K754E (SEQ ID NO:62) predicted using nnPredict analysis is shown below in TABLE 69.

TABLE 69



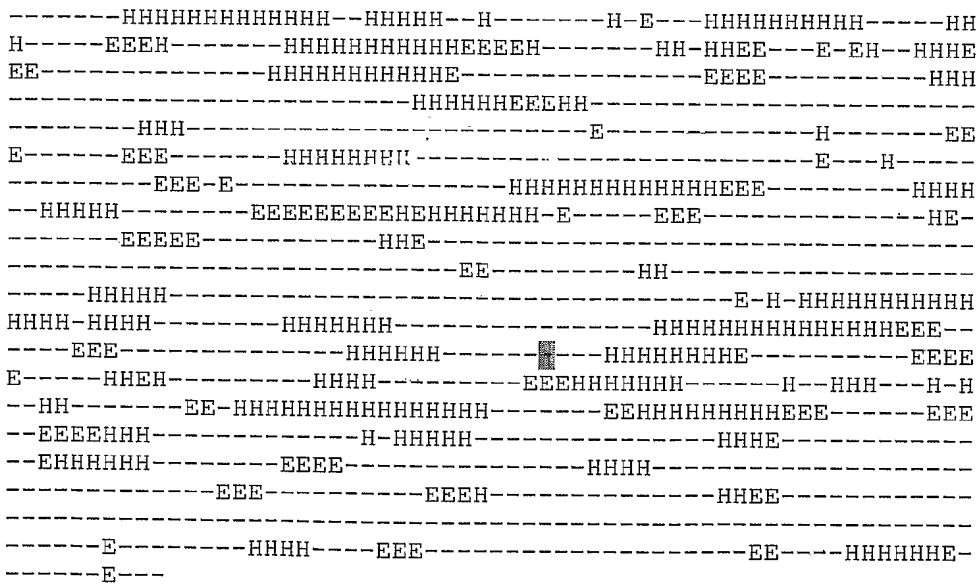
[251] The amino acid sequence of EGFR mutant polypeptide K754R (SEQ ID NO:63) is shown below in TABLE 70. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 70

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSP**R**ANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:63)

[252] A schematic representation of the secondary structure of EGFR mutant polypeptide K754R (SEQ ID NO:63) predicted using nnPredict analysis is shown below in TABLE 71.

TABLE 71



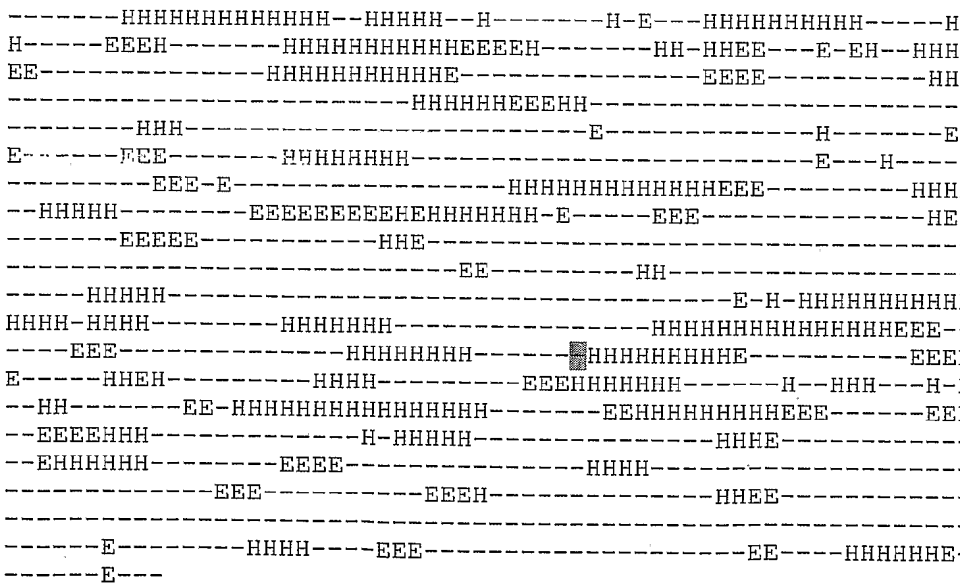
[253] The amino acid sequence of EGFR mutant polypeptide N756S (SEQ ID NO:64) is shown below in TABLE 72. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 72

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKA**s**KEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:64)

[254] A schematic representation of the secondary structure of EGFR mutant polypeptide N756S (SEQ ID NO:64) predicted using nnPredict analysis is shown below in TABLE 73.

TABLE 73



[255] The amino acid sequence of EGFR mutant polypeptide K757R (SEQ ID NO:65) is shown below in TABLE 74. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 74

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKAN**r**EILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:65)

[256] A schematic representation of the secondary structure of EGFR mutant polypeptide K757R (SEQ ID NO:65) predicted using nnPredict analysis is shown below in TABLE 75.

TABLE 75

```
-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE--------------HHHHHHHHHHE---------------EEEE----------HHH
------------------------HHHHHEEEHH----------------------
--------HHH-------------------------E-------------H-------EE
E------EEE-------HHHHHHHH-------------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHHHHHHH-E-----EEE-------------HE-
-------EEEEE-----------HHE---------------------------------
-----------------------EE---------HH-------------------
-----HHHHH-----------------------------E-H-HHHHHHHHHH
HHHH-HHHH--------HHHHHHH---------------HHHHHHHHHHHHHHEEE--
----EEE--------------HHHHHHHH-------HHHHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHH-------EEHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE----------
--EHHHHHH--------EEEE---------------HHHH-----------------
-------------EEE----------EEEH-------------HHEE----------
-----------------------------------------------------
------E--------HHHH----EEE-----------------EE----HHHHHHE-
------E---
```

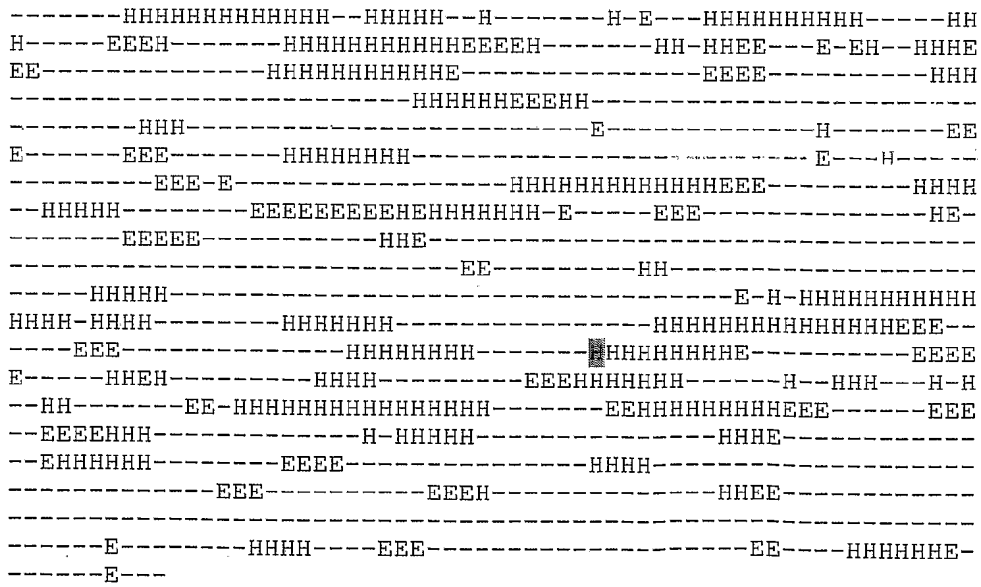
[257] The amino acid sequence of EGFR mutant polypeptide K757E (SEQ ID NO:66) is shown below in TABLE 76. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 76

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKAN**e**EILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:66)

[258] A schematic representation of the secondary structure of EGFR mutant polypeptide K757E (SEQ ID NO:66) predicted using nnPredict analysis is shown below in TABLE 77.

TABLE 77



[259] The amino acid sequence of EGFR mutant polypeptide E758G (SEQ ID NO:67) is shown below in TABLE 78. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 78

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANK**G**ILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:67)

[260] A schematic representation of the secondary structure of EGFR mutant polypeptide E758G (SEQ ID NO:67) predicted using nnPredict analysis is shown below in TABLE 79.

TABLE 79

```
-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE--------------HHHHHHHHHHHE---------------EEEE----------HHH
------------------------HHHHHHEEEHH----------------------
--------HHH-------------------------E-----------H-------EE
E------EEE-------HHHHHHHH----------------------E---H-----
---------EEE-E----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE-------------HE-
-------EEEEE----------HHE-----------------------------
----------------------------EE---------HH-----------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHH
HHHH-HHHH--------HHHHHHH---------------HHHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHH--------▓-HHHHHHE----------EEEE
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE----------
--EHHHHHH--------EEEE---------------HHHH-----------------
-------------EEE----------EEEH-------------HHEE----------
------------------------------------------------------------
------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---
```

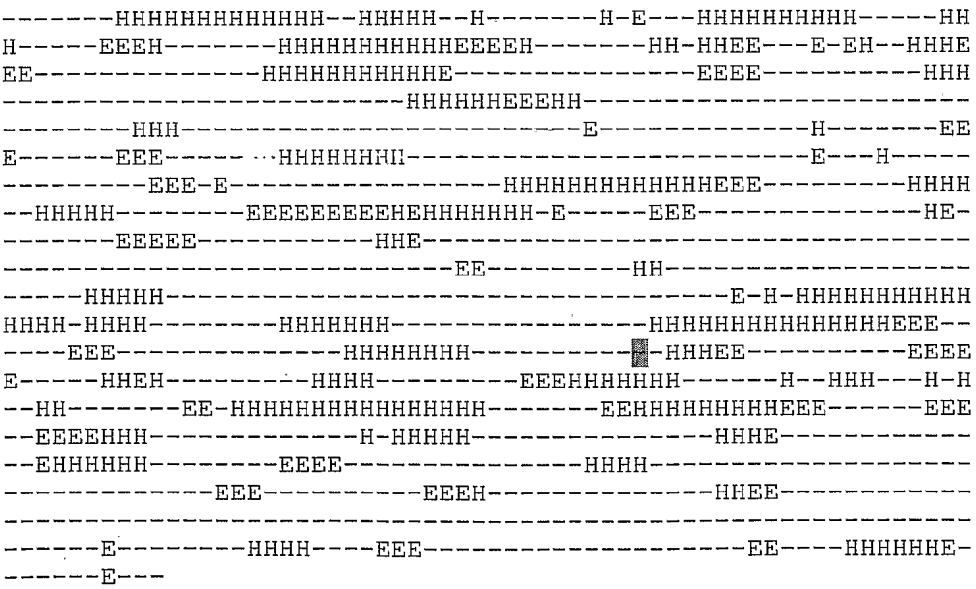
[261] The amino acid sequence of EGFR mutant polypeptide L760P (SEQ ID NO:68) is shown below in TABLE 80. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 80

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEI**p**DEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:68)

[262] A schematic representation of the secondary structure of EGFR mutant polypeptide L760P (SEQ ID NO:68) predicted using nnPredict analysis is shown below in TABLE 81.

TABLE 81



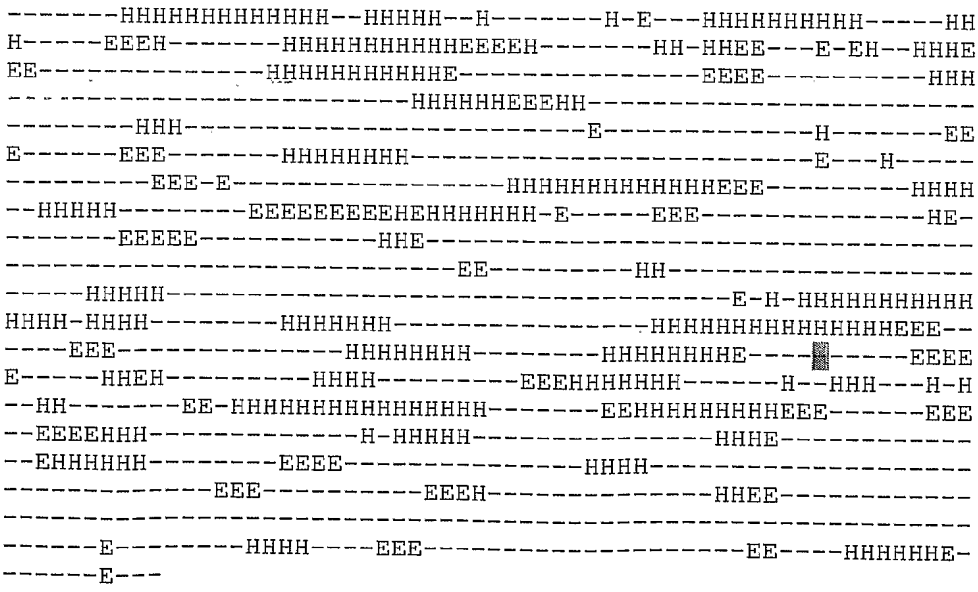
[263] The amino acid sequence of EGFR mutant polypeptide N771D (SEQ ID NO:69) is shown below in TABLE 82. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 82

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVD**D**PHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:69)

[264] A schematic representation of the secondary structure of EGFR mutant polypeptide N771DG221W (SEQ ID NO:69) predicted using nnPredict analysis is shown below in TABLE 83.

TABLE 83



[265] The amino acid sequence of EGFR mutant polypeptide V774M (SEQ ID NO:70) is shown below in TABLE 84. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 84

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHMCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:70)

[266] A schematic representation of the secondary structure of EGFR mutant polypeptide V774M (SEQ ID NO:70) predicted using nnPredict analysis is shown below in TABLE 85.

TABLE 85

-------HHHHHHHHHHHHH--HHHHH--H-------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE--------------HHHHHHHHHHHE---------------EEEE----------HHH
------------------------HHHHHHEEEHH-----------------------
--------HHH-----------------------E-------------H-------EE
E------EEE-------HHHHHHHH-------------------------- E-- -- --
---------EEE-E-----------------HHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE--------------HE-
-------EEEEE-----------HHE---------------------------------
----------------------------EE---------HH-------------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH--------HHHHHHH----------------HHHHHHHHHHHHHHHEEE--
----EEE-------------HHHHHHHH--------HHHHHHHHE-------H-HE--E
E-----HHEH---------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHEEE------EEE
--EEEEHHH-------------H-HHHHH---------------HHHE------------
--EHHHHHH--------EEEE---------------HHHH--------------------
-------------EEE----------EEEH--------------HHEE------------

------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---

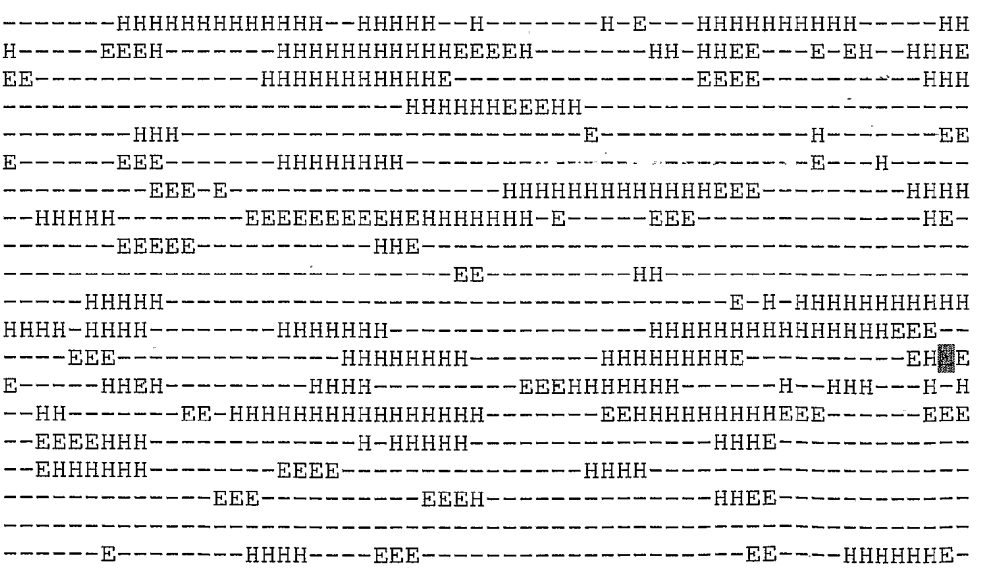
[267] The amino acid sequence of EGFR mutant polypeptide G779S (SEQ ID NO:71) is shown below in TABLE 86. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 86

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLL**s**I
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:71)

[268] A schematic representation of the secondary structure of EGFR mutant polypeptide G779S (SEQ ID NO:71) predicted using nnPredict analysis is shown below in TABLE 87.

TABLE 87



[269] The amino acid sequence of EGFR mutant polypeptide I890T (SEQ ID NO:72) is shown below in TABLE 88. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 88

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRTYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:72)

[270] A schematic representation of the secondary structure of EGFR mutant polypeptide I890T (SEQ ID NO:89) predicted using nnPredict analysis is shown below in TABLE 89.

TABLE 89

-------HHHHHHHHHHHHH--HHHHH--H--------H-E---HHHHHHHHHH-----HH
H-----EEEH-------HHHHHHHHHHHHEEEEH-------HH-HHEE---E-EH--HHHE
EE--------------HHHHHHHHHHHE--------------EEEE----------HHH
------------------------HHHHHHEEEHH----------------------
--------HHH----------------------E-------------H-------EE
E------EEE------HHHHHHHH--------------------E---H-----
---------EEE-E-----------------HHHHHHHHHHHHHHEEE---------HHHH
--HHHHH--------EEEEEEEEEHEHHHHHHH-E-----EEE--------------HE-
-------EEEEE----------HHE----------------------------------
---------------------------EE---------HH--------------------
-----HHHHH-----------------------------------E-H-HHHHHHHHHHH
HHHH-HHHH--------HHHHHHH----------------HHHHHHHHHHHHHHHEEE--
----EEE--------------HHHHHHHH--------HHHHHHHHE----------EEEE
E-----HHEH----------HHHH---------EEEHHHHHHH------H--HHH---H-H
--HH-------EE-HHHHHHHHHHHHHHHH-------EEHHHHHHHHHH-------EEE
--EEEEHHH-------------H-HHHHH----------------HHHE----------
--EHHHHHH--------EEEE-------------HHHH----------------------
-------------EEE----------EEEH-------------HHEE------------

------E--------HHHH----EEE-------------------EE----HHHHHHE-
------E---

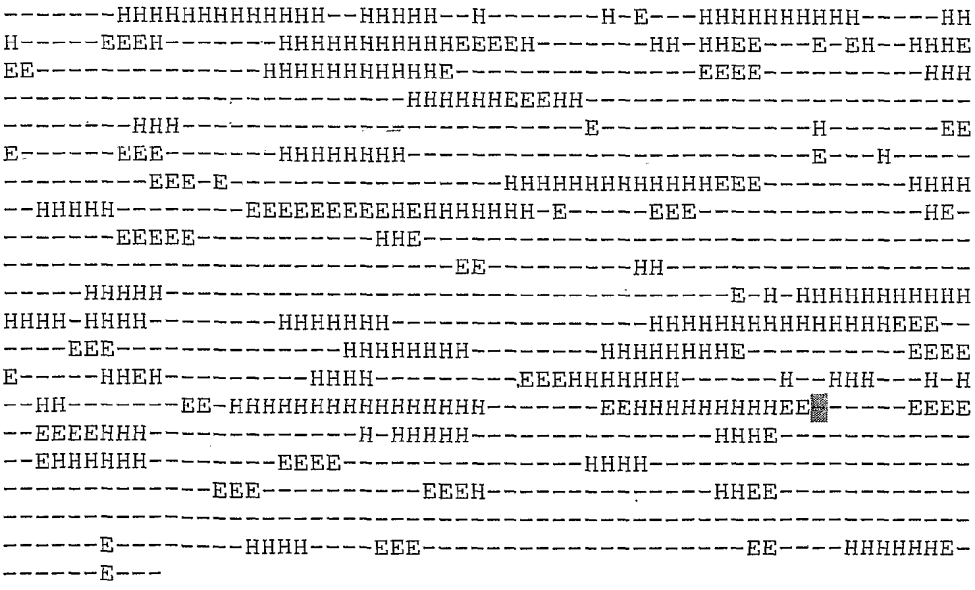
[271] The amino acid sequence of EGFR mutant polypeptide Y891S (SEQ ID NO:73) is shown below in TABLE 86. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 90

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRI**S**THQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:73)

[272] A schematic representation of the secondary structure of EGFR mutant polypeptide Y891S (SEQ ID NO:73) predicted using nnPredict analysis is shown below in TABLE 91.

TABLE 91



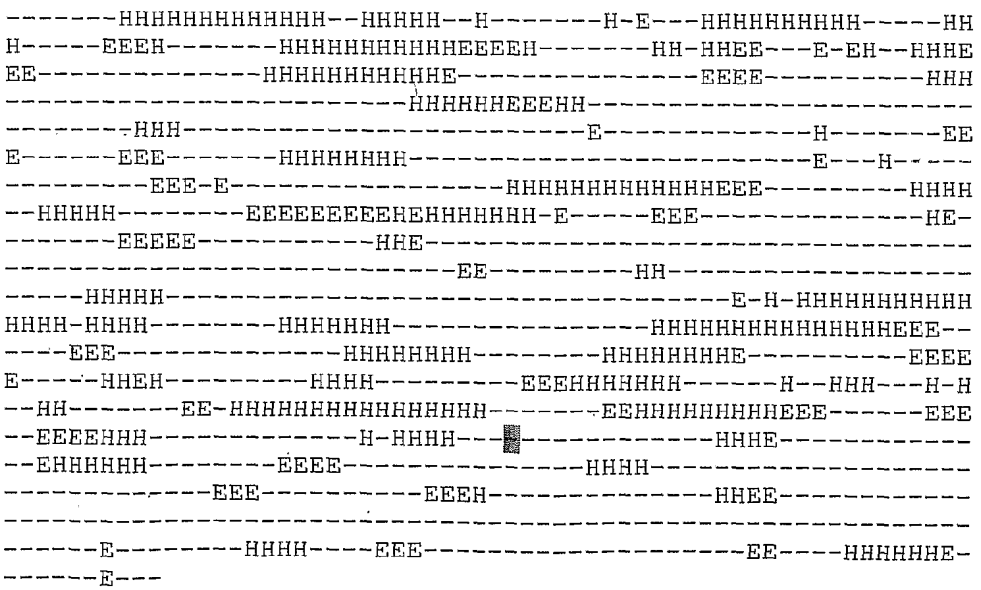
[273] The amino acid sequence of EGFR mutant polypeptide R932G (SEQ ID NO:74) is shown below in TABLE 92. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 92

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGE**G**LPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPS**P**TDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:74)

[274] A schematic representation of the secondary structure of EGFR mutant polypeptide R932G (SEQ ID NO:74) predicted using nnPredict analysis is shown below in TABLE 93.

TABLE 93



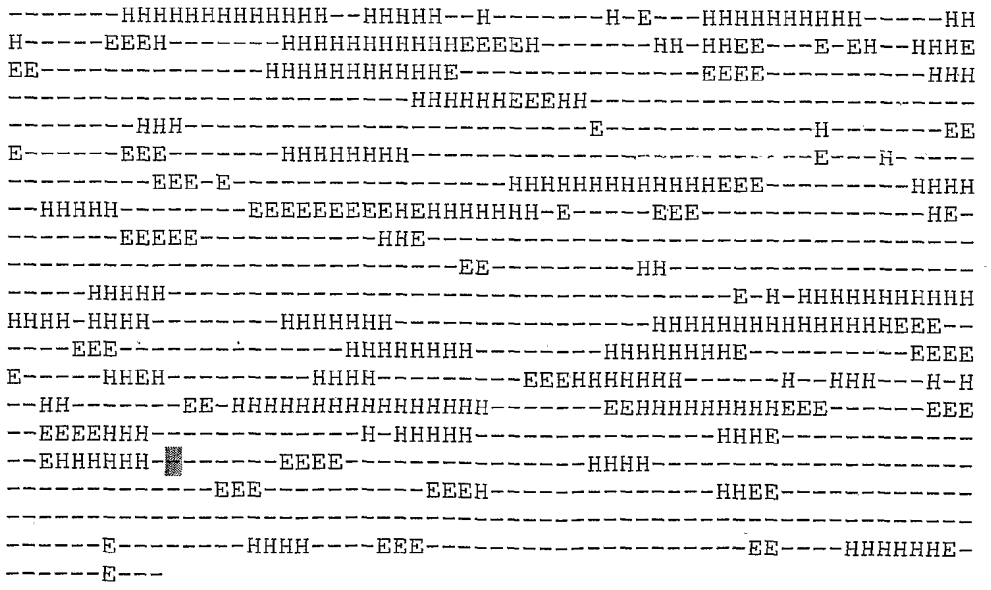
[275] The amino acid sequence of EGFR mutant polypeptide M971R (SEQ ID NO:75) is shown below in TABLE 94. The position of the mutated amino acid residue is highlighted as a lower case character in bold underlined text.

TABLE 94

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKRARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA (SEQ ID NO:75)

[276] A schematic representation of the secondary structure of EGFR mutant polypeptide M971R (SEQ ID NO:75) predicted using nnPredict analysis is shown below in TABLE 95.

TABLE 95



[277] The influence of EGFR mutations on EGFR protein secondary structure is summarized below in TABLE 92.

TABLE 92
Influence of EGFR mutations on EGFR protein secondary structure

Mutation	Predicted Protein Secondary Structure Change
Q217R	This mutation eliminates the upstream short helix and converts the protein secondary structure to a strand orientation
G221W	No change is predicted
G239C	No change is predicted
C251F	This mutation extends a helix, changes the secondary structure of 2 amino acids
E282D	No change is predicted
L509V	This mutation change the secondary structure of 2 amino acids
R521K	No change is predicted
D587N	No change is predicted
C595W	No change is predicted
H618Y	This mutation introduces a 3-amino acid strand
T693I	No change is predicted
P699A	This mutation extends a downstream helix, changes the secondary structure of 3 amino acids
E709V	No change is predicted
T725P	This mutation eliminates a strand, changes the secondary structure of 3 amino acids
Y727S	This mutation shortens upstream strand by 1 amino acid, changes the secondary structure of 2 amino acids
G729S	This mutation change the secondary structure of 4 amino acids
I732T	No change is predicted
G735R	No change is predicted
E736D	This mutation changes the secondary structure of 2 amino acids
V738G	This mutation changes the secondary structure of 1amino acid
I740T	No change is predicted
K745R	This mutation changes the secondary structure of 1 amino acid
R748G	This mutation changes the secondary structure of 4 amino acids
R748I	This mutation extends a helix by 1 amino acid
T751I	This mutation extends the upstream helix by 2 amino acids
T751A	This mutation extends the upstream helix by 1 amino acid
K754E	This mutation shortens the upstream helix by 2 amino acids
K754R	This mutation shortens the upstream helix by 2 amino acids
N756S	This mutation extends the downstream helix by 1 amino acid
Y891S	This mutation changes the secondary structure of 1 amino acid
K757E	This mutation extends the downstream helix by 1 amino acid
K757R	This mutation extends the downstream helix by 1 amino acid
E758G	This mutation shortens the downstream helix by 2 amino acids

TABLE 92
Influence of EGFR mutations on EGFR protein secondary structure

Mutation	Predicted Protein Secondary Structure Change
L760P	This mutation shortens the downstream helix by 4 amino acids, change the secondary structure of 5 amino acids in total
N771D	No change is predicted
V774M	This mutation change the secondary structure of 4 amino acids
G779S	This mutation change the secondary structure of 2 amino acids
I890T	No change is predicted
Y891S	This mutation change the secondary structure of 1 amino acid
Q894Term	This mutation loses the C-terminal
Y900Term	This mutation loses the C-terminal
R932G	No change is predicted
M971R	No change is predicted

[278] Among the forty-three distinct mutations identified for EGFR, L760P appears to have most impact on the protein structure. According to the 3D crystal structure of the protein kinase domain of the wild-type human EGFR, L760 is located in the middle of a alpha-helix of about ten amino acids. Mutation of leucine to proline at 760 can break the alpha-helix as indicated by the secondary structure analysis. Other mutations at positions located within the protein kinase domain (between 712 and 968), *e.g.*, I732T, G735R, K754R, T751I and N756S etc., can also impact EGFR function by altering its structure. In addition, mutations at positions E282, T693, T725, T751, K754, N756, L760, and Y891 may alter the phosphorylation patterns of the mutated and nearby sites. For example, T751 is predicted as a potential threonine phosphorylation site that is next a potential serine phosphorylation site at 752. Amino acid alterations at positions 751 (T751I & T751A), 754 (K754R & K754E) and 756 (N756S) can alter the phosphorylation pattern and consequently the biological activity of the EGFR polypeptide. Mutations G729R, K745R, G779S and R932G are identified at highly conserved positions. Q217R, G725R, M971R, N771D, H618Y, R748G, R748I, R932G, D587N, E709V, E758G, K754E and K757E changes the number of electric charges of EGFR.

[279] A summary of the results of computational analysis of the effect of the EGFR missense mutations identified in the present invention on select features of wild-type EGFR is provided below in TABLE 97.

TABLE 97
Evaluation of Mutations by Sequence Features

Mutation	Protein domain	Phospho-rylation	Other modification	AA conservation	AA property change	Secondary Structure
Q217R					+	+
G221W						
G239C				+		
C251F				+		+
L267V						
E282D		++				
D587N				+	+	
L509V						+
C595W				+		
H618Y					+	+
T693I		+++				
P699A						+
E709V					+	
T725P		++	+			+
Y727S			+			+
G729R	+++		+	+		++
I732T			+			
G735R			+	+	+	
E736D			+			+
V738G			+	+		+
I740T			+	+		
K745R	+++		+			+
R748G					+	++

+: the effect of mutation on protein function is low
++: the effect of mutation on protein function is medium
+++: the effect of mutation on protein function is high

TABLE 97
Evaluation of Mutations by Sequence Features

Mutation	Protein domain	Phospho-rylation	Other modification	AA conservation	AA property change	Secondary Structure
R748I					+	+
T751I		++		+		+
T751A		++		+		+
K754E		++			++	+
K754R		++				+
N756S		++				+
K757R						+
K757E					++	+
E758G				+	+	+
L760P		++				++
N771D					+	
V774M						++
G779S	+++					+
I890T						
Y891S		++				+
Q894Term						+++
Y900Term						+++
R932G	+++				+	
M971R					+	

+: the effect of mutation on protein function is low
++: the effect of mutation on protein function is medium
+++: the effect of mutation on protein function is high

[280] Among the forty-three distinct mutations identified for EGFR, L760P appears to have most impact on the protein structure. According to the 3D crystal structure of the protein kinase domain of the wild-type human EGFR, L760 is located in the middle of a alpha-helix of about ten amino acids. Mutation of leucine to proline at 760 can break the alpha-helix as indicated by the secondary structure analysis. Other mutations at positions located within the protein kinase domain (between 712 and 968), *e.g.*, I732T, G735R, K754R, T751I and N756S etc., can also impact EGFR function by altering its structure. In addition, mutations at positions E282, T693, T725, T751, K754, N756, L760, and Y891 may alter the phosphorylation patterns of the mutated and nearby sites. For example, T751 is predicted as a potential threonine phosphorylation site that is next a potential serine phosphorylation site at 752. Amino acid alterations at positions 751 (T751I & T751A), 754 (K754R & K754E) and 756 (N756S) can alter the phosphorylation pattern and consequently the biological activity of the EGFR polypeptide. Mutations G729R, K745R, G779S and R932G are identified at highly

conserved positions. Q217R, G725R, M971R, N771D, H618Y, R748G, R748I, R932G, D587N, E709V, E758G, K754E and K757E changes the number of electric charges of EGFR.

EXAMPLE 3

ANALYSIS OF EGFR MUTATION FOR THERANOSTIC CANCER TREATMENT IN A SUBJECT

[281] An agent that modulates EGFR biological activity (*e.g.*, EGFR antagonist) is administered to a patient with cancer, *e.g.*, glioblastoma, breast cancer, cholangioma, non-small-cell lung cancer (NSCLC), melanoma, ovarian cancer, prostate cancer, colon cancer and myeloma, when the patient has a SNP/mutation pattern that correlates with the disease. In one embodiment, the SNPs and mutations are selected from the group consisting the EGFR mutations and polymorphisms summarized in TABLE 1.

[282] In a preferred embodiment, the EGFR antagonist is AEE788, which inhibits multiple receptor tyrosine kinases including EGFR, HER2, and VEGFR, to stimulate tumour cell growth and angiogenesis. Traxler P *et al.*, *Cancer Res.* 64(14): 4931-4941 (July 15, 2004). In preclinical studies, AEE788 showed high target specificity and demonstrated antiproliferative effects against tumour cell lines and in animal models of cancer. AEE788 also exhibited direct antiangiogenic activity. AEE788 is currently in phase I clinical development.

[283] In another embodiment, the EGFR antagonist is gefitinib (Iressa®), which has been approved by the Food and Drug Administration (FDA) as a single agent for the treatment of non-small cell lung cancer (NSCLC) that has progressed after, or failed to respond to two other types of chemotherapy (drugs used to kill cancer cells). Iressa® belongs to a group of anticancer drugs called epidermal growth factor receptor-tyrosine kinase inhibitors (EGFR-TKI). Iressa® is given by mouth as a single tablet of 250 mg with or without food. Patients with poorly tolerated diarrhoea (sometimes associated with dehydration) or skin drug reactions may be successfully managed by providing a brief (up to 14 days) therapy interruption followed by starting again with the 250 mg daily dose.

[284] In yet another embodiment, the EGFR antagonist is erlotinib (Tarceva®; OSI-774). Erlotinib is an epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor used in molecular targeted therapy. Erlotinib is used to treat non-small cell lung cancer. It is also being studied in many other cancers including breast cancer. Erlotinib is a pill taken by

mouth each day as directed with a large glass of water, at least one hour before or two hours after a meal. For lung cancer patients, the recommended starting dose is 150 mg each day.

[285] Other EGFR-targeting agents include PKI166 (Fabbro D *et al., Pharmacol Ther.* 93(2-3):79-98 (February-March 2002); Traxler P et al., Med. Res. Rev. 21(6):499-512 (November 2001)), C-225, ZD1839.

EQUIVALENTS

[286] The present invention is not to be limited in terms of the particular embodiments described in this application, which are intended as single illustrations of individual aspects of the invention. Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. Functionally equivalent methods and apparatuses within the scope of the invention, in addition to those enumerated herein, will be apparent to those skilled in the art from the foregoing descriptions. Such modifications and variations are intended to fall within the scope of the appended claims. The present invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

CLAIMS

We claim:

1. The use of an EGFR modulating agent in the manufacture of a medicament for the treatment of cancer in a selected patient population, wherein the patient population is selected on the basis of the genotype of the patients at a EGFR genetic locus indicative of a propensity for having cancer.

2. The use of an EGFR modulating agent according to claim 1, wherein the EGFR modulating agent is selected from the group consisting of AEE788 and PKI166.

3. The use of an EGFR modulating agent according to claim 1, wherein the cancer is selected from the group consisting of: glioblastoma; melanoma; ovarian cancer; breast cancer; cholangioma; non-small-cell lung cancer (NSCLC); prostate cancer; colon cancer; and myeloma.

4. An isolated polynucleotide having a sequence encoding an EGFR mutation listed in TABLE 1.

5. A vector comprising a polynucleotide of claim 4.

6. An organism containing a polynucleotide of claim 4.

7. The polynucleotide of claim 4, further comprising a polynucleotide sequence encoding an EGFR polypeptide having a sequence selected from the group consisting of: SEQ ID NO:36 - SEQ ID NO:75.

8. An isolated polypeptide having a sequence selected from the group consisting of SEQ ID NO:36 - SEQ ID NO:75.

9. A method for treating cancer in a subject, comprising the steps of:
   (a) obtaining the genotype or haplotype of a subject at a EGFR gene locus, wherein the genotype and/or haplotype is indicative of a propensity for having cancer; and
   (b) administering an anti-cancer therapy to the subject.

10. The method of claim 9, wherein the anti-cancer therapy is selected from the group consisting of Glivec®, FEMARA®, Sandostatin® LAR® , ZOMETA®, vatalanib, everolimus, gimatecan, patupilone, midostaurin, pasireotide, LBH589, AEE788 and AMN107.

11. The method of claim 9, wherein the cancer is selected from the group consisting of: glioblastoma; breast cancer; melanoma, ovarian cancer, cholangioma; non-small-cell lung cancer (NSCLC); prostate cancer; colon cancer; and myeloma.

12. The method of claim 9, wherein the genotype is heterozygous, with at least one of the alleles containing an EGFR polymorphism and/or mutation of TABLE 1.

13. The method of claim 9, wherein the genotype is homozygous, with at least one of the alleles containing an EGFR mutation or polymorphism of TABLE 1.

14. The method of claim 9, wherein the anti-cancer therapy is the administration of a therapeutically effective amount of an EGFR modulating agent.

15. A method for diagnosing a propensity for having cancer in a subject, comprising the steps of:
   (a) obtaining the genotype or haplotype of a subject at a EGFR gene locus, wherein the genotype and/or haplotype is indicative of a propensity of the cancer to respond to the drug; and
   (b) identifying the subject as having a propensity for having cancer.

16. A method for choosing subjects for inclusion in a clinical trial for determining efficacy of an EGFR modulating agent, comprising the steps of:
    (a) interrogating the genotype and/or haplotype of a subject at an EGFR gene locus;
    (b) then:
        (i) including the subject in the trial if the genotype is indicative of a propensity to cancer by the subject;
        (ii) excluding the subject from the trial if the genotype is not indicative of a propensity to cancer by the subject; or
        (iii) both (i) and (ii).

17. The method of claim 16, wherein the cancer is selected from the group consisting of: glioblastoma; breast cancer; melanoma; ovarian cancer; cholangioma; non-small-cell lung cancer (NSCLC); prostate cancer; colon cancer; and myeloma.

18. A kit for use in determining a treatment strategy for cancer, comprising:
    (a) a reagent for detecting a polynucleotide encoding an EGFR mutation and/or polymorphism of TABLE 1;
    (b) a container for the reagent; and
    (c) a written product on, or in, the container describing the use of the polynucleotide in determining a treatment strategy for the cancer.

19. The kit of claim 18, wherein the cancer is selected from the group consisting of: glioblastoma; breast cancer; melanoma; ovarian cancer; cholangioma; non-small-cell lung cancer (NSCLC); prostate cancer; colon cancer; and myeloma.

20. The kit of claim 18, wherein the reagent for detecting the polynucleotide encoding an EGFR mutation of TABLE 1 is a set of primer pairs that hybridize to a polynucleotide on either side of the EGFR mutations and polymorphisms of TABLE 1.

21. The use of a polynucleotide having a sequence encoding an EGFR mutation listed in TABLE 1 as a drug target.

22. An antibody that binds to a polypeptide having a sequence selected from the group consisting of SEQ ID NOS:36-75.